CLEI LACLO 2018

CLATG

I CONCURSO LATINO-AMERICANO DE TRABALHOS DE GRADUAÇÃO

v. 1 | 2018

Organizadores
Maria Amelia Eliseo
Ana Claudia Rossi

SBC
Sociedade Brasileira de Computação

Organização

Fomento

Organizadores
Ana Claudia Rossi
Maria Amelia Eliseo

CLATG - I CONCURSO LATINO-AMERICANO DE TRABALHOS DE GRADUAÇÃO

1ª edição

Porto Alegre

Sociedade Brasileira de Computação - SBC

2018

C744

CLATG – I Concurso latino-americano de trabalhos de graduação [livro eletrônico] / Organizadores Maria Amelia Eliseo, Ana Claudia Rossi – Porto Alegre: Sociedade Brasileira de Computação, 2018.
3.853 Kb ; PDF.

Anais do CLATG – I Concurso latino-americano de trabalhos de graduação.
ISBN: 978-85-7669-459-5.
Livro eletrônico.

1. Sistemas computacionais. 2. Infraestrutura de rede. 3. Inteligência artificial. 4. Interação humano-computador. 5. Informática na educação. 6. Educação em computação. 7. Gestão de dados e informações. 8. Processamento de imagem I. Eliseo, Maria Amelia. II. Rossi, Ana Claudia. III.Título.

CDD 005

Bibliotecária Responsável: Marta Luciane Toyoda – CRB 8/ 8234

I Concurso Latino-americano de Trabalhos de Graduação

01 e 02 de outubro de 2018. São Paulo, São Paulo, Brasil.
Evento satélite ao CLEI-LACLO 2018 (XLIV Conferência Latino-americana de Informática e XIII Conferência Latino-americana de Tecnologias de Aprendizagem).

Comitê Organizador

Ana Claudia Rossi (Mackenzie)
Maria Amélia Eliseo (Mackenzie)

Comitê Científico

Ana Claudia Rossi (Mackenzie)
Ana Grasielle Corrêa (Mackenzie)
Anarosa Alves Franco Brandão (USP)
Andre Riyuiti Hirakawa (USP)
André Luiz Satoshi Kawamoto (UTFPR)
Calebe de Paula Bianchini (Mackenzie)
Cesar Alberto Collazos (Universidad del Cauca)
Daniela Vieira Cunha (Mackenzie)
Denise Stringhini (UNIFESP)
Diego Roberto Colombo Dias (UFSJ)
Dirceu Matheus Junior (Mackenzie)
Fernando Figueira Filho (UFRN)
Jorge Luís Risco Becerra (USP)
Leandro Pupo (Mackenzie)
Manuel Ibarra Cabrera (Universidad Nacional Micaela Bastidas de Apurímac)
Marcelo de Paiva Guimarães (UNIFESP)
Maria Amelia Eliseo (Mackenzie)
Maurício Marengoni (Mackenzie)
Paula Peres (Politécnico do Porto)
Pedro Paulo B. Oliveira (Mackenzie)
Pollyana Notargiacomo (Mackenzie)
Romero Tori (USP)
Sara Paiva (Instituto Politécnico de Viana do Castelo)
Solange Nice Alves de Souza (USP)
Valéria Farinazzo Martins (Mackenzie)

Comissão Julgadora

Prof. Dr. Jorge Risco (USP)
Prof. Dr. Paulo Lopes (Mackenzie)
Prof. Dr. Paulo Sampaio (UNIFACS)
Profª. Drª. Anarosa A. F. Brandão (USP)

Prefácio

O CLATG - Concurso Latino-Americano de Trabalhos de Graduação dedica-se à apresentação e discussão de trabalhos concluídos por estudantes no término de seu curso de graduação em computação (Ciência da Computação, Engenharia de Computação, Sistemas de Informação, Engenharia de Software e afins). Durante o evento, os egressos da graduação têm a chance de apresentar seus trabalhos num ambiente construtivo e propício para discussões, de forma a trocar ideias e conhecimento com pesquisadores experientes e profissionais da área. É uma oportunidade de mostrar à comunidade científica o potencial dos trabalhos desenvolvidos na graduação e incentivar a continuidade da pesquisa numa eventual pós-graduação ou inserção no mercado.

O I CLATG foi um evento satélite realizado no contexto dos eventos CLEI 2018 (XLIV Conferência Latino-americana de Informática) e LACLO 2018 (XIII Conferência Latino-americana de Tecnologias de Aprendizagem), ocorridos em conjunto na cidade de São Paulo. Os trabalhos enviados ao I CLATG foram avaliados por um Comitê Científico, sendo aceitos 38% dos artigos enviados. Os trabalhos aceitos, um total de 11, foram apresentados em sessões de apresentação oral e avaliados por uma Comissão Julgadora formada por pesquisadores na área da computação. Foram selecionados os três melhores trabalhos, sendo o primeiro lugar premiado.

Esta edição do livro “CLATG - I CONCURSO LATINO-AMERICANO DE TRABALHOS DE GRADUAÇÃO” reúne os textos de todos os trabalhos aceitos e apresentados no I CLATG. Desta forma, colabora com a divulgação da produção acadêmica ao transmitir o trabalho dos estudantes latino-americanos que recentemente completaram sua graduação.

Agradecemos à comunidade pelo interesse em enviar seus trabalhos e compartilhar conhecimentos e, em especial, todos aqueles que contribuíram na edição deste livro, incluindo a Comissão Julgadora e o Comitê Científico (Comitê de Programa).

Maria Amelia Eliseo
Ana Claudia Rossi

CLATG

Sumário

1. Redes Ópticas Elásticas

Chapter

1

Enrutamiento, Nivel de Modulación y Asignación de Espectro en Redes Ópticas Elásticas Una comparación entre enfoques de solución basados en ILP

Ivan Ríos y Diego P. Pinto-Roa

Abstract

This work presents the impact of the k-shortest path approach on the spectrum used, and the joint and serial optimization approaches in the Routing, Modulation Level and Spectrum Allocation problem (RMLSA). In this context, the 4 models of integer linear programming (ILP) are implemented under the considerations of joint/serial optimization approach and path/link oriented routing approach. In general, the performance of path-oriented routing approaches tend to that of link-oriented routing approaches, however, these latter approaches require a longer computation time to achieve substantial improvements, which makes path-oriented routing approaches more attractive when the problem becomes more complex.

Resumo

Este trabajo presenta el impacto del enfoque k-shortest path sobre el espectro utilizado, y los enfoques de optimización conjunta y serializada, en el problema de Enrutamiento, Nivel de Modulación y Asignación de Espectro (RMLSA). En este contexto, se implementan los 4 modelos de programación lineal entera (ILP) bajo las consideraciones de enfoque de optimización conjunta/serial y de enrutamiento orientado a camino/enlace. En general, el desempeño de los enfoques de enrutamiento orientado a camino tiende al de los enfoques de enrutamiento orientado a enlace. Sin embargo, estos últimos enfoques necesitan un mayor tiempo de cómputo para lograr mejoras substanciales, lo cual hace que los enfoques de enrutamiento orientado a camino sean más atractivos cuando el problema se vuelve más complejo.

1.1. Introducción

La red óptica elástica (*Elastic Optical Networks*, EON) [Jinno et al., 2009] es una red con capacidad de asignación de espectro a los caminos ópticos de acuerdo a los requerimientos de ancho de banda de los clientes. El espectro está dividido en reducidas ranuras de frecuencia (*Frequency Slots*, FS) donde cada una de las conexiones ópticas tiene asignado un diferente número de ranuras consecutivas. Para realizar un tráfico, la red EON impone restricciones de capa óptica [Chatterjee et al., 2015]: (i) continuidad, (ii) contigüidad, y (iii) no solapamiento del espectro asignado a las solicitudes. Bajo estas condiciones el problema del Enrutamiento y la Asignación del Espectro (*Routing and Spectrum Allocation*, RSA) es el principal desafío en EON [Delvalle et al., 2016]. Dependiendo del formato de modulación de la señal óptica transmitida se puede disminuir el espectro asignado. El formato de modulación se encuentra asociado a la distancia entre el nodo fuente y destino. Si la distancia no supera 500 Km, 1000 Km y 2000 Km se asigna la modulación 8QAM, QPSK y BPSK, respectivamente [Li and Kim, 2015]. En este contexto el problema RSA clásico se extiende al problema de Ruteo y Asignación de Espectro y nivel de Modulación (*Routing, Modulation Level and Spectrum Allocation*, RMLSA) [Klinkowski et al., 2011, Christodoulopoulos et al., 2011]. En este trabajo se estudia el efecto del número de K rutas más cortas sobre el espectro utilizado, y los enfoques de optimización conjunta y serializada, en el problema de Enrutamiento, Nivel de Modulación y Asignación de Espectro (RMLSA), siendo estos análisis los principales aportes de este trabajo considerando soluciones exactas basadas en Programación Lineal Entera (*Integer Linear Programming*, ILP).

1.2. Redes Ópticas Elásticas

La creciente demanda de Internet requiere una infraestructura escalable y eficiente en costo [Christodoulopoulos et al., 2011] y han hecho que los operadores de red busquen soluciones que permitan mayor capacidad y flexibilidad en sus redes ópticas, las cuales utilizan la tecnología de transmisión denominada Multiplexación por División de Longitud de onda (*Wavelength Division Multiplexing*, WDM) [Mayoral López de Lerma et al., 2013]. WDM se clasifica en dos tipos: WDM Densa (*Dense WDM*, DWDM) y WDM Ligera (*Coarse WDM*, CWDM) [Ramaswami, 2006]. Desafortunadamente, las conexiones de 400 Gbps no podrán implementarse en ranuras de 50 GHz que son las más comúnmente utilizadas [Gerstel et al., 2012]. Por esta razón, se debería aumentar la separación espectral, por ejemplo a 100 GHz. Pero se estaría desperdiciando una gran cantidad de espectro para solicitudes de conexión menores de 100 Gbps [Ahumada et al., 2014].

Para solucionar este inconveniente, Jinno et al. presentaron las Redes Ópticas Elásticas (*Elastic Optical Networks*, EON) [Jinno et al., 2009]. Con esta propuesta, EON logra un uso más eficiente del espectro con respecto a las redes tradicionales basadas en WDM [Chatterjee et al., 2015].

1.2.1. El problema RSA

El problema RSA en EON es análogo al problema de Enrutamiento y Asignación de Longitud de Onda (*Routing and Wavelength Assignment*, RWA) [Rodas-Brítez and Pinto-Roa, 2014] en redes basadas en WDM [Pinto-Roa et al., 2011]. La diferencia entre RSA

y RWA es la capacidad de asignar de manera flexible y más granular el espectro de frecuencias [Liu and Gu, 2012]. En RSA, un conjunto de FS, es asignado a una conexión en vez de una longitud de onda en RWA. Dependiendo del tipo de tráfico el problema RSA se puede clasificar en: *Offline* o Estático y *Online* o Dinámico [Castro et al., 2012]. En el caso estático ya se tiene de antemano las solicitudes de conexión, en cambio, en el caso dinámico, las solicitudes de conexión van llegando en el tiempo.
El planteamiento más simple del problema RSA se basa en una demanda dada, compuesta por un par origen-destino y una cantidad de FS requeridos, en la que se busca encontrar una ruta que tenga disponible al menos, la cantidad de FS solicitados y respetando las restricciones de capa óptica explicadas en el párrafo anterior.

1.2.2. Asignación de Nivel de Modulación

En el problema RSA no se aborda el problema de la modulación de la señal. En [Christodoulopoulos et al., 2011] se propone considerar distintos esquemas de modulación para mejorar el uso del espectro. En este contexto, una vez obtenida la longitud l del camino elegido se aborda el problema de asignación de nivel de modulación (*Modulation Level*, ML) para dicha ruta. La idea fundamental es asignar una modulación que maximice la tasa de datos de transmisión de forma a minimizar el número de FS utilizados [Christodoulopoulos et al., 2011]. Cuanto más corto sea el camino elegido se utilizará el nivel de modulación con mayor tasa de transmisión. En este trabajo, se asume que los formatos de modulación BPSK, QPSK, y 8QAM, son los utilizados para transmisiones de hasta $l_1 \leq$ 2000 Km, $l_2 \leq 1000$ Km y $l_3 \leq 500$ Km, respectivamente (como se muestra en la Figura 1.1). Posterior al cálculo del nivel de modulación se procede a calcular el número de FS óptimo. Se asume λ como la demanda solicitada, $R \in \{1,2,3\}$ el nivel de modulación óptimo asociado a la ruta de longitud l, y C el ancho de banda para un FS con modulación de un bit por símbolo (BPSK), entonces la cantidad de FS mínima que puede ser asignado a la solicitud es $\lceil \lambda / \{R \cdot C\} \rceil$. Por ejemplo, si $\lambda = 5 \cdot C$ y la longitud de la ruta es $l = 400$ Km, entonces corresponde a una modulación 8QAM (R_3) con espectro igual a $\lceil 5 \cdot C \ / 3 \cdot C \rceil =$ 2 ranuras de frecuencia. En cambio, si la ruta tiene una longitud de entre 500 Km y 1000 Km se necesitan 3 ranuras de frecuencia para modulación QPSK. Finalmente, si la ruta tiene una longitud de entre 1000 Km y 2000 Km se necesitan 5 ranuras de frecuencia con modulación BPSK.

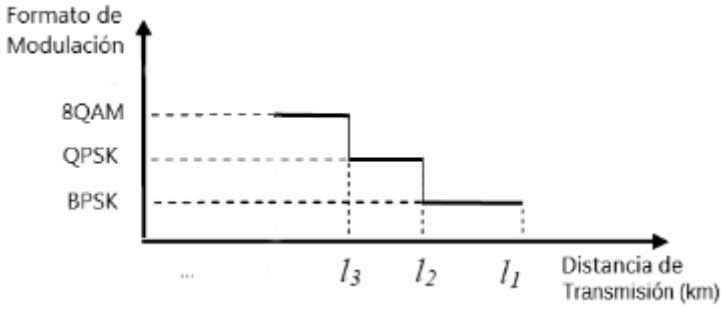

Figure 1.1. Formato de modulación versus Distancia de transmisión [Christodoulopoulos et al., 2011]

1.2.3. Problema RMLSA

El problema RSA considerando el problema de nivel de modulación (ML) se denomina RMLSA, el cual se puede plantear de la siguiente forma. Dada una topología de una EON, y una matriz de demanda de espectro con modulación base BPSK, se debe asignar una ruta física, el nivel de modulación y el espectro a cada demanda con el fin de minimizar el máximo espectro utilizado en cualquier enlace en la red, bajo las restricciones de capa óptica. El problema RMLSA es el foco del presente trabajo.

1.3. Trabajos relacionados en RMLSA

La literatura reporta varios enfoques de solución, de los cuales nos centramos en los basados en ILP (la búsqueda de trabajos relacionados se realizó con la herramienta denominada Google Académico con las palabras claves RMLSA e ILP). Básicamente, las estrategias que abordan el problema RMLSA con ILP se pueden clasificar según el enfoque de optimización y de enrutamiento:

1. enfoque de optimización (*optimization approach*):
   (a) optimización serial (*serial optimization*): en este enfoque se resuelven los problemas RML y SA en 2 fases. En la primera fase se resuelve el problema RML y en la segunda fase se resuelve el problema SA. Este enfoque ha sido propuesto en [Christodoulopoulos et al., 2011] y usualmente la literatura lo indica como RML+SA.
   (b) optimización conjunta (*joint optimization*): en este enfoque se resuelven los problemas RML y SA de manera simultánea [Christodoulopoulos et al., 2011, Li and Kim, 2015] y usualmente se lo indica como RMLSA.
2. enfoque de enrutamiento (*routing approach*):
   (a) enrutamiento orientado a camino (*path-oriented*): en este enfoque se considera que una heurística calcula k caminos candidatos sobre los cuales el optimizador de enrutamiento escoge el mejor camino acorde al criterio de optimización [Christodoulopoulos et al., 2011].
   (b) enrutamiento orientado a enlace (*link-oriented*): en este enfoque se consideran todos los caminos posibles de la red por lo que todos los enlaces son candidatos potenciales para formar una ruta [Li and Kim, 2015].

El problema RMLSA ha sido tratado en ampliamente en la literatura. En el 2011, fue inicialmente abordado por [Christodoulopoulos et al., 2011], donde se agrega el nivel de modulación al problema RSA y se proponen dos modelos de ILP basados en ruta para el problema RMLSA.

En el 2016, [Li and Kim, 2015] presenta un modelo de ILP basado en enlaces para resolver el problema RMLSA en EON con la restricción de alcance de transmisión óptica.

Se puede notar que ningún trabajo ha propuesto un esquema que considere un enfoque de optimización serial y enrutamiento orientado a enlaces simultáneamente. Con el objeto llevar a cabo el estudio propuesto en este trabajo, se contribuye al estado del

arte un enfoque de optimización serial y enrutamiento orientado a enlaces inspirado en un modelo ILP [Li and Kim, 2015].

En este trabajo, se propone el modelo ILP de [Li and Kim, 2015] dividido en 2 fases. Es importante el estudio que se realiza en este paper de manera a saber qué tan efectivos y escalables pueden ser los algoritmos propuestos en la literatura, y ver en cada caso qué tan aproximada está la solución de cada algoritmo a la solución óptima.

1.4. Pruebas experimentales

1.4.1. Ambiente Experimental

Los modelos ILP de la literatura (JP, SP y JL) y el modelo ILP propuesto (SL). fueron implementados en IBM ILOG CPLEX Optimization Studio Versión 12.6, y la automatización del experimento en JAVA en una computadora con procesador Intel Xeon Processor E5530 Quad Core CPU (2.40 GHz) y RAM de 15 GB.

1.4.2. Esquema y Escenarios Experimentales

Se realizaron 2 tipos de pruebas: pruebas con carga uniforme y pruebas con carga aleatoria.
Las pruebas uniformes se realizaron una sola vez por cada petición de $\lambda = \{100, 150, 200, 250, 300, 350, 400\}$ FS (basado en el experimento realizado en [Li and Kim, 2015]), y considerando los valores $k = 1, 2, 3, 4, 5$ para el k-shortesth path. El tiempo de optimización para la carga uniforme fue fijado como máximo 4 horas. El framework fue implementado en JAVA versión 7 mientras que los modelos fueron implementados en CPLEX versión 12.6.0.0.
En las pruebas aleatorias se consideraron 10 instancias para cada caso (100-150, 150-200, 200-250, 250-300, 300-350 y 350-400 FS) generando solicitudes de todos los nodos contra todos con valores comprendidos en los rangos mencionados. La forma de ejecución se realizó con un generador de números aleatorios y con un script se realizaron las ejecuciones de los 4 algoritmos con valor de $k = 5$ rutas. El tiempo de optimización para la carga aleatoria fue fijado como máximo 1 hora. Los parámetros que se han utilizado en estas ejecuciones fueron $F_{total} = 100000$ y $GB = 1$.

La topología utilizada como escenario experimental fue una porción de la red genérica DT (extraída de [Li and Kim, 2015]) con distancias reales entre los nodos, como se muestra en la Figura 1.2 con solicitudes de todos los nodos contra todos.

1.4.3. Resultado y Discusión

1.4.3.1. Carga uniforme

La Figura 1.3 indica el espectro calculado para el modelo SP, mientras que la Figura 1.4 indica el espectro calculado para el modelo JP. Note que SP y JP son enfoques orientados a camino.

Observando la Figura 1.3 se ha notado que a partir del valor $k = 2$ disminuye considerablemente el espectro utilizado, con el valor $k = 3$ no se reduce aún más el espectro pero todavía la solución no converge a la solución óptima, con el valor $k = 4$ se obtiene

Figure 1.2. Topología de red genérica DT parcial, con 10 nodos y 16 enlaces no dirigidos (extraída de [Li and Kim, 2015])

nuevamente una mejoría del resultado, llegando a la solución cercana a la óptima, y finalmente, con el valor $k = 5$ ya no se observa ningún cambio, por lo cual, para este modelo con la topología DT parcial el valor $k = 4$ es el valor mínimo.

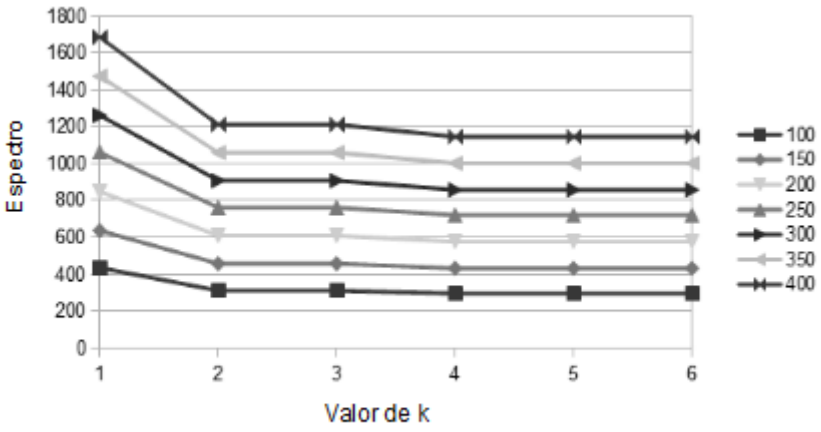

Figure 1.3. Espectro para distintas cargas de tráfico para SP

Observando la Figura 1.4 se ha notado que a partir del valor $k = 2$ disminuye también considerablemente el espectro utilizado. Con el valor $k = 3$ el espectro se mantiene a excepción de la carga de tráfico de 200 ranuras de frecuencia. Con el valor $k = 4$ se obtiene en la mayoría de los casos una mejoría del resultado, llegando a la solución cercana a la óptima. Finalmente, con el valor $k = 5$ solamente se ha notado un cambio para la carga de

tráfico de 250 ranuras, por lo cual, para este modelo con la topología DT parcial el valor $k = 5$ es el valor mínimo.

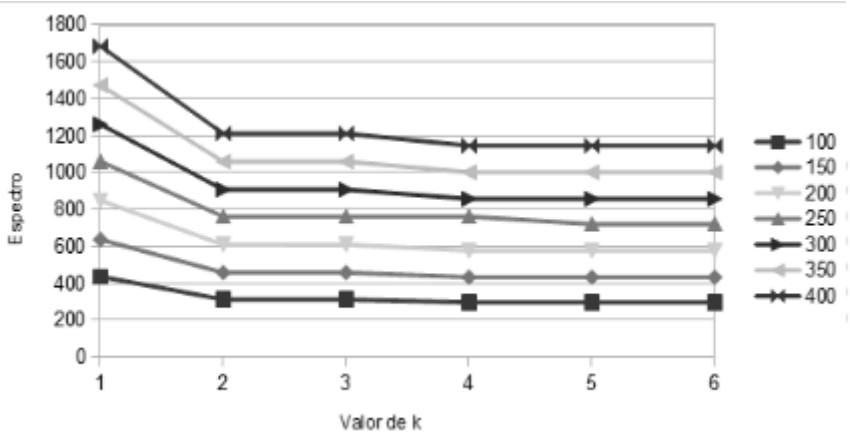

Figure 1.4. Espectro para distintas cargas de tráfico para JP

Relación de costos para enfoques basados en enlaces

Observando la Figura 1.5 se ha notado que para todas las cargas de tráfico el modelo SL tiene siempre soluciones iguales o mejores que las del modelo JL, debido a que el modelo JL no logra converger a la solución óptima debido a que el tiempo de cómputo máximo configurado es de 4 horas. El modelo JL necesita mayor tiempo de cómputo, dado que opera en un mayor espacio de búsqueda.

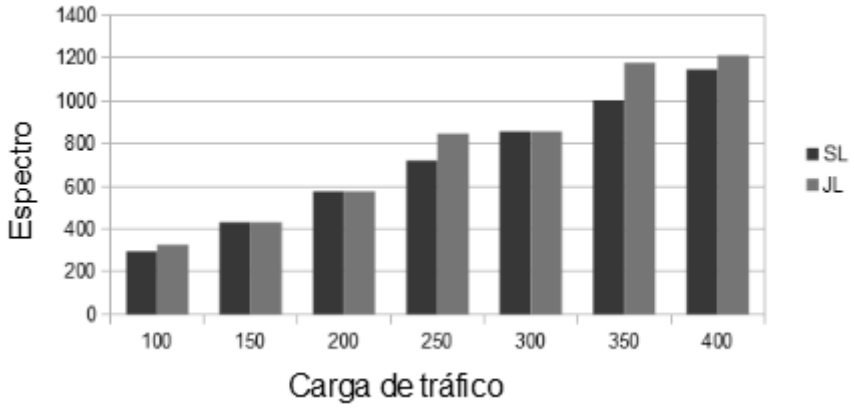

Figure 1.5. Espectro obtenido por los modelos ILP para 10 ejecuciones con cargas aleatorias entre 100 y 150 ranuras de frecuencia

1.5. Carga aleatoria

Los experimentos fueron realizados con el valor $k = 5$.

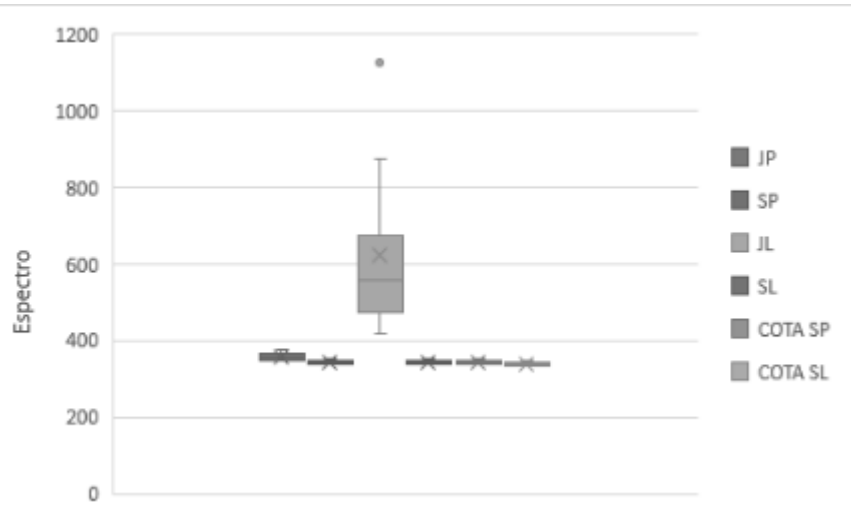

Figure 1.6. Espectro obtenido por los modelos ILP para 10 ejecuciones con cargas aleatorias entre 100 y 150 ranuras de frecuencia

Observando la Figura 1.6 se ha notado que el modelo JL tiene una mayor dispersión de las soluciones y está muy por arriba el espectro calculado de los demás modelos. El modelo JP solo está un poco más arriba de los otros modelos en cuanto a los rangos de soluciones. Las cotas inferiores calculadas son los valores ideales que podría tener la cantidad de FS utilizados en el mejor de los casos previa a la asignación de espectro. Esto se observa en las columnas COTA SP y COTA SL. Los valores de la COTA SL son menores o iguales a los valores de la COTA SP. Las cotas calculadas corresponden a c_{RML} de los enfoques orientados a camino y a enlace, respectivamente. Se ha notado que los valores de las cotas son menores o iguales que los valores de los modelos correspondientes. Los valores que arrojan los modelos SP y SL son iguales en la mayoría de las ejecuciones.

1.6. Conclusiones

Al aumentar el número de rutas K el espectro disminuye, y aumenta el tiempo de cómputo. Entre los enfoques de optimización serial y conjunta conviene mucho más el enfoque serial debido a que se logra converger a una solución sub-óptima en poco tiempo. El enfoque conjunto no converge a la solución óptima en muchos casos. Entre los enfoques orientados a camino y a enlaces el último enfoque no necesita definir la cantidad de caminos K para calcular el espectro máximo, sin embargo, los enfoques orientados a camino son mucho más rápidos en tiempo de ejecución.

Referencias

[Ahumada et al., 2014] Ahumada, R., Lopez, A. L., Villalobos, F. A., Massmann, S. F., and Castro, G. F. (2014). Spectrum allocation algorithms for elastic DWDM networks on dynamic operation. *IEEE Latin America Transactions*, 12(6):1012–1018.

[Castro et al., 2012] Castro, A., Velasco, L., Ruiz, M., Klinkowski, M., Fernandez-Palacios, J., and Careglio, D. (2012). Dynamic Routing and Spectrum (Re) Allocation in Future Flexgrid Optical Networks. *Computer Networks*, 56(12):2869–2883.

[Chatterjee et al., 2015] Chatterjee, B. C., Sarma, N., and Oki, E. (2015). Routing and spectrum allocation in elastic optical networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 17(3):1776–1800.

[Christodoulopoulos et al., 2011] Christodoulopoulos, K., Tomkos, I., and Varvarigos, E. (2011). Elastic Bandwidth Allocation in Flexible OFDM-based Optical Networks. *IEEE/OSA Journal of Lightwave Technology*, 29(9):1354–1366.

[Delvalle et al., 2016] Delvalle, L., Alfonzo, E., and Roa, D. P. P. (2016). EONS: An online RSA simulator for elastic optical networks. In *Computer Science Society (SCCC), 2016 35th International Conference of the Chilean*, pages 1–12. IEEE.

[Gerstel et al., 2012] Gerstel, O., Jinno, M., Lord, A., and Ben Yoo, S. (2012). Elastic Optical Networking: A New Dawn for the Optical Layer? *IEEE Communications Magazine*, 50(2):s12–s20.

[Jinno et al., 2009] Jinno, M., Takara, H., Kozicki, B., Tsukishima, Y., Sone, Y., and Matsuoka, S. (2009). Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. *IEEE Communications Magazine*, 47(11).

[Klinkowski et al., 2011] Klinkowski, M., Walkowiak, K., and Jaworski, M. (2011). Off-line algorithms for routing, modulation level, and spectrum assignment in elastic optical networks. In *Transparent Optical Networks (ICTON), 2011 13th International Conference on*, pages 1–6. IEEE.

[Li and Kim, 2015] Li, B. and Kim, Y.-C. (2015). Efficient routing and spectrum allocation considering QoT in elastic optical networks. *IEEE Communications Surveys & Tutorials*, pages 109–112.

[Liu and Gu, 2012] Liu, J. and Gu, W. (2012). Spectrum consecutiveness based routing and spectrum allocation in flexible bandwidth networks. *Chinese Optics Letters*.

[Mayoral López de Lerma et al., 2013] Mayoral López de Lerma, A. et al. (2013). Algoritmos de planificación para redes elásticas. B.S. thesis.

[Pinto-Roa et al., 2011] Pinto-Roa, D. P., Barán, B., and Brizuela, C. A. (2011). Routing and wavelength converter allocation in WDM networks: a multi-objective evolutionary optimization approach. *Photonic Network Communications*, 22(1):23–45.

[Ramaswami, 2006] Ramaswami, R. (2006). Optical networking technologies: what worked and what didn't. *IEEE Communications Magazine*, 44(9):132–139.

[Rodas-Brítez and Pinto-Roa, 2014] Rodas-Brítez, M. D. and Pinto-Roa, D. P. (2014). Quality of protection on WDM networks: A quantitative paradigm based on recovery probability. In *Computing Conference (CLEI), 2014 XL Latin American*, pages 1–11. IEEE.

2. Desenvolvimento de Jogos Digitais

Chapter

2

Desenvolvimento de Jogos Digitais Educacionais para Incentivo ao Processo de Ensino Aprendizagem

Elias Sousa Leitão e Carla Marina Costa Paxiúba

Resumo

Os jogos digitais têm a capacidade de prender a atenção, entreter e estimular, sendo mecanismos que podem ser usados também para ensinar. Assim, este trabalho tem por objetivo descrever, modelar, desenvolver e aplicar jogos digitais educacionais que possam ser adotados por professores e estudantes como ferramentas que proporcionem entretenimento, diversão e auxilie na fixação de conceitos do curriculum escolar. Foram desenvolvidos dois jogos digitais, o primeiro tem o objetivo de dar assistência ao processo de Alfabetização infantil e o segundo de auxiliar a ministração e/ou fixação dos conteúdos relacionados à Geometria Plana. Deste trabalho resulta-se a criação de ferramentas que podem contribuir com instituições de ensino, auxiliando professores e pais interessados em fazer uso de métodos que visem favorecer o ensino e aprendizagem.

1. Introdução

As tecnologias de informação e comunicação (TIC) têm o poder de modificar as interações sociais e, cada vez mais, tornam-se parte do cotidiano das pessoas, principalmente das crianças e adolescentes que convivem diariamente com os recursos tecnológicos [Potapczuk, 2013].

As TIC, entre elas os jogos digitais, têm o potencial para serem utilizadas como ferramentas mediadoras no processo de ensino-aprendizagem de alunos [Potapczuk, 2013]. Nisto, os jogos digitais têm a capacidade de prender a atenção, entreter e estimular nos jovens um grande fascínio, funcionando como mecanismos que podem ser usados também para ensiná-los. Neste sentido, os jogos podem ser utilizados para fortalecimento do processo de aprendizagem, desde que promovam o desenvolvimento de capacidades e estratégias importantes para o crescimento intelectual e cognitivo dos estudantes [Vargas, 2016].

Assim, tendo como base a importância dos jogos para o ensino e aprendizagem, este trabalho tem por objetivo descrever, modelar, desenvolver e aplicar jogos digitais educacionais que possam ser adotados por professores e/ou estudantes, como ferramentas que proporcionem entretenimento, diversão e auxilie na fixação de conceitos do curriculum escolar. Os jogos aqui apresentados têm por intuito servir para assistência ao processo de Alfabetização e Letramento infantil e para auxiliar a

ministração e/ou fixação dos conteúdos relacionados às figuras geométricas básicas da Geometria Plana.

2. Referencial Teórico

Impulsionado pelas tecnologias, o jogo digital pode funcionar como ferramenta dinâmica capaz de unificar o conhecimento e a estratégia no ambiente da sala de aula. Como instrumento de aprendizagem, o jogo digital educacional é inovador, pois integra conteúdos programáticos e estratégias para alcançar um objetivo educacional específico. Os jogos educacionais são elaborados para prender a atenção do jogador ao disponibilizar desafios que exigem crescente nível de habilidades, estimulando autoaprendizagem, descoberta e curiosidade por meio de desafios [Menezes, Muzatti, & Taquaritinga, 2016].

Neste contexto, todas as disciplinas do quadro escolar podem ser favorecidas com a utilização das mídias digitais, basta para tanto a criação e/ou adequação destas para os fins educacionais.

A alfabetização infantil é uma das temáticas que podem ser beneficiadas pelos games, pois podem despertar na criança o interesse associado à descoberta de algo novo. Os elementos que compõem os jogos digitais, como: os personagens; os cenários; as cores; os sons; e os desafios intrínsecos a eles, vinculados a conteúdos educativos, podem estimular no aluno sua curiosidade, de modo que esta ultrapasse os limites da sala de aula, levando-o a buscar o conhecimento por sua própria vontade.

A aprendizagem por meio de metodologias que envolvem elementos visuais e sonoros é mais atrativa e torna-se mais fácil. Segundo Silva et al. (2008), algumas mídias digitais, dentre elas os jogos, podem possibilitar ao estudante a exploração, investigação e descoberta das informações, estimulando a construção de seu próprio conhecimento.

Para que o processo de Alfabetização seja bem estruturado, de modo a desempenhar um papel significativo para a aprendizagem, ele deve estar vinculado a objetos de interesse e significativos para o indivíduo. Neste sentido, os jogos têm uma função muito importante pois proporcionam a interação da criança com algo de seu maior interesse, o ato de brincar [Pizarro, 2012].

De forma semelhante, outra disciplina que pode usufruir da utilização dos jogos, para a divulgação de seus saberes, é a Geometria. Ela tem grande relevância dentro do cenário escolar, pois nasce da compreensão de como os objetos podem ser representados conceitualmente.

Com base na sua importância, a utilização de jogos para o ensino da Geometria possibilita ao aluno a construção de seu conhecimento, dando-lhe o poder de manipular, investigar, criar estratégias para a resolução de problemas e assimilação de conceitos geométricos, além de despertar-lhe o prazer pela aprendizagem [Barros, 2012]. Sendo assim, faz-se relevante buscar mecanismos que reforcem o ensino da Geometria, tornando-a mais atrativa aos olhos do estudante, motivando-o a aprender e deixando-o mais propenso a desenvolver seu potencial [Lima, 2014].

Neste sentido, o segundo jogo digital proposto por este trabalho é voltado ao ensino da Geometria, e na sua temática desenvolve os conceitos ministrados durante o I Ciclo de Formação do Ensino Fundamental (1º ao 3º ano).

3. Desenvolvimento

Esta seção apresenta a especificação e o desenvolvimento dos jogos digitais frutos deste trabalho, descrevendo suas funcionalidades. Os jogos foram projetados observando-se as ideias apresentadas anteriormente, mostrando as especificações e os requisitos de ambos os jogos destinados à Alfabetização e ao ensino de Geometria.

3.1. Ensino de Geometria: Jogo "Busca Geométrica"

O jogo digital destinado ao ensino da Geometria foi o primeiro a ser desenvolvido e nomeia-se como "Busca Geométrica". Trata-se de um jogo educacional bidimensional (2D) estilo Plataforma1, o qual tem por objetivo principal auxiliar a ministração e/ou fixação dos conteúdos relacionados às formas geométricas básicas da Geometria Plana.

No ambiente de game, o usuário representado por um personagem animado tem por objetivo percorrer o cenário do jogo recolhendo os objetos que lhe forem apresentados, no caso, as formas geométricas. Cada um dos elementos ao serem coletados mostram suas características e informações. O jogador deverá também completar os desafios de cada fase e, consequentemente, conseguir pontuação necessária para completar as etapas do jogo.

3.1.1. Personagem Principal

O personagem principal (Figura 1) percorre os cenários do jogo, recolhendo as formas geométricas e é usado como ponte para a visualização das informações e características das formas geométricas. Ele deverá completar os desafios e a pontuação necessária para completar as fases do jogo.

Figura 1. Personagem principal do jogo que é controlado pelo usuário

3.1.2. Personagens Vilões

Os personagens vilões estão divididos em duas categorias, os vilões terrestres e aéreos. Os vilões têm o objetivo de abater o personagem principal durante o seu trajeto no ambiente do game, agindo como obstáculos que devem ser superados.

Os vilões terrestres são pequenas gosmas, chamadas de GloomyBads (Figura 2). São diferentes personagens, cada um representado por uma cor e comportamento específico. Retiram todos os pontos do jogador quando tocados.

[1] Jogo de plataforma é um gênero de jogos eletrônicos em que o jogador corre e pula entre plataformas e obstáculos, enfrentando inimigos e coletando objetos prêmios.

De modo semelhante, há os vilões aéreos (voadores) nomeados como SkyBads (Figura 2). Cada um deles possui tamanho e velocidade própria, lançam bombas ou espinhos que retiram pontos do jogador.

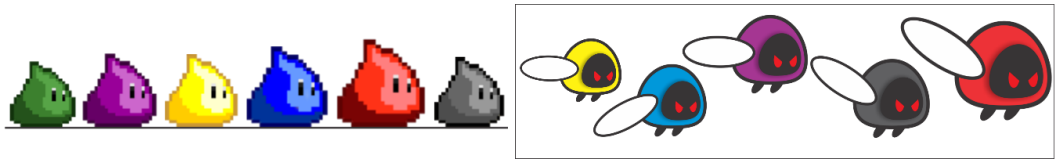

Figura 2. Personagens vilões do jogo, à esquerda são apresentados os GloomyBads, à direita os SkyBads

3.1.3. Plataformas

As plataformas representam parte do cenário do jogo e são utilizadas como base para a movimentação dos personagens. Os elementos são montados no cenário similarmente a ideia da montagem das peças de um quebra-cabeça. Estes elementos (Figura 3) foram adquiridos gratuitamente no site *opengameart.org*[2].

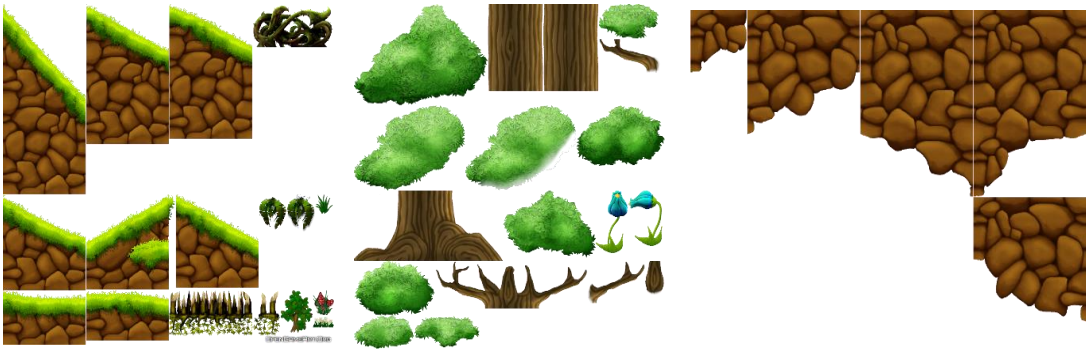

Figura 3. Elementos de montagem utilizados para a construção do cenário e das plataformas.

3.1.4. Formas Geométricas

As formas geométricas (Figura 4) são um dos elementos mais importantes do jogo, pois simbolizam seu real objetivo: a apresentação das características das formas geométricas da Geometria Plana. A interação delas com o jogador ocorre quando o personagem colidi com estes elementos, onde lhe é apresentado uma mensagem com as informações pertinentes a forma geométrica coletada.

Mensagens, com informações das características das formas geométricas, são mostradas ao usuário todas as vezes que este encontrar uma figura geométrica nova ou quando lhe for apresentado um novo desafio para ser completado.

[2]Endereço completo: <http://opengameart.org/content/forest-themed-sprites>.

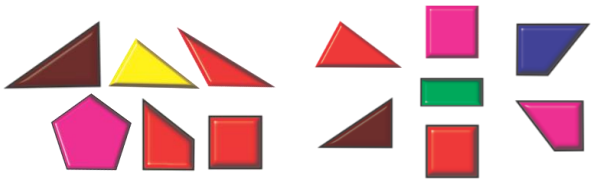

Figura 4. Elementos geométricos que são utilizados no jogo.

3.1.5. Execução e Premissas do Jogo Busca Geométrica

O jogo, ao ser executado, mostra ao usuário sua tela inicial e após clicar no botão central lhe será apresentada uma subtela com uma história de contextualização para o jogo. Em seguida, o jogo entra em execução.

O usuário poderá movimentar o personagem utilizando as teclas direcionais do teclado (setas: esquerda; direita; para cima). O jogador deverá percorrer o cenário, onde lhe será apresentado as primeiras informações do jogo e o que deve ser feito para avançar.

Quando o jogo se inicia um cronômetro é ativado para indicar o tempo da partida. Em cena é mostrada ainda a quantidade e os tipos de figuras geométricas que são coletadas pelo jogador, como observado na Figura 5. A cada fase o usuário disporá de cinco "vidas", representadas por cinco corações na parte superior da tela. Estas vidas poderão ser perdidas, caso o jogador seja pego nas armadilhas dos vilões sem ter em sua conta alguma das formas geométricas ou caso ele morra ao colidir com os obstáculos.

O jogo é composto por duas fases. Na 1ª fase o personagem terá contato somente com triângulos. São também propostos ao usuário desafios que devem ser completados até o termino da fase. Na 1ª fase é apresentada ao usuário uma "estrela" formada apenas por triângulos (Figura 5), esta deverá ser inteiramente preenchida até ao término da fase. Sempre que o usuário conseguir um novo triangulo diferente, ele apresentará suas características e será encaixado adequadamente à estrela. Após completar os desafios da 1ª fase, o jogador terá que responder um quis de questões relacionadas às formas geométricas que lhe foram apresentadas. São questões semelhantes a demonstrada na Figura 6. Com os conhecimentos que adquiriu no jogo, deverá indicar qual alternativa satisfaz o problema.

Figura 5. Jogo em execução. Na 1ª fase o jogador deverá completar a estrela com os triângulos que encontrar durante seu trajeto

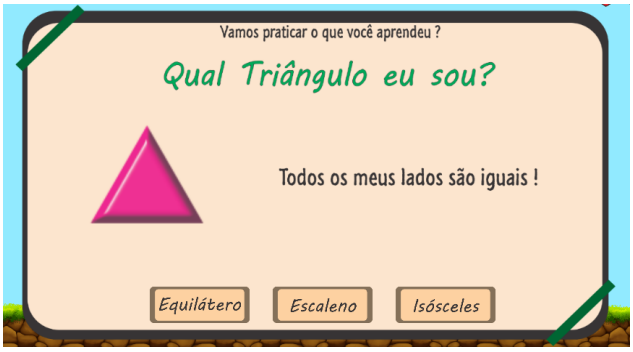

Figura 6. Exemplo do questionário aplicado ao final de cada fase do jogo

A 2ª fase do jogo obedece a mesma sequência lógica, porém nesta serão apresentadas novas figuras geométricas, como quadrados, retângulos, trapézios e pentágonos. Nesta fase há um novo desafio, o jogador deverá coletar todas as figuras geométricas que dão forma a um "coração" (Figura 7). Ao final da fase, o usuário terá novamente que responder questões que são contextualizadas nas características das formas geométricas presentes no decorrer desta nova etapa.

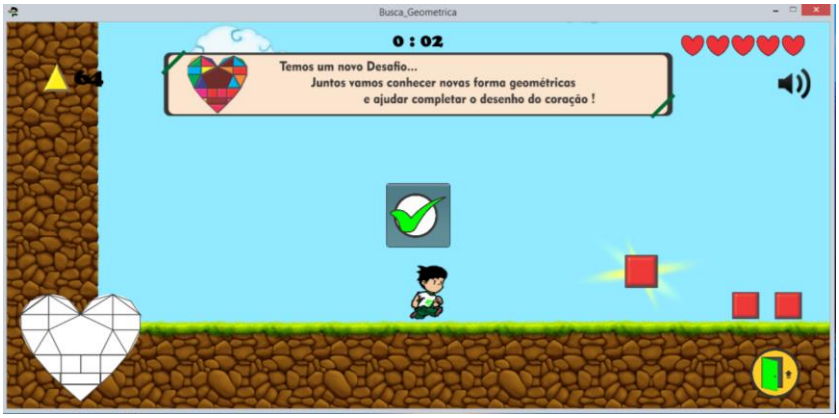

Figura 7. Apresentação do desafio para a 2ª fase do jogo, o jogador terá que completar o coração formado por figuras geométricas

3.2. Processo de Alfabetização: Jogo "Alfabeto Voador"

O segundo jogo desenvolvido foi nomeado como "Alfabeto Voador". Trata-se de um jogo educacional em 2D que dá suporte nos primeiros contatos da criança com o alfabeto, no reconhecimento de sílabas e na construção de palavras. Neste jogo, o usuário terá o controle do personagem principal, representado por um pequeno avião animado. Basicamente o jogador terá que controlar o avião horizontalmente para que possa capturar as letras ou sílabas que caem em sua direção. Os elementos devem ser coletados obedecendo uma ordem pré-estabelecida. A cada objeto capturado pelo jogador, em sua sequência correta, é reproduzido o som específico da pronúncia do elemento alfabético capturado e o jogador ganha pontos.

3.2.1. Personagem Principal

Um Avião (Figura 8) representa o personagem principal do jogo, tem por função recolhendo os objetos em cena que caem verticalmente em sua direção. Deve recolher os elementos alfabéticos e completar o desafio de cada etapa.

Figura 8. Personagem principal que é controlado pelo jogador

3.2.2. Elementos do Alfabeto

Os elementos coletados pelo Avião são as letras e sílabas do alfabeto, sendo as letras também usadas para a construção das palavras na fase posterior. Cada vez que um elemento for coletado na ordem correta, o jogador ganha pontos, o elemento da sequência é marcado como coletado e o som da pronúncia de tal letra ou sílaba coletados é reproduzido. Na Figura 9 são exemplificados os elementos de alfabetização que são utilizados no jogo.

Figura 9. Representação dos elementos alfabéticos apresentados no jogo.

3.2.3. Execução e Premissas do Jogo

Quando o jogo é executado é apresento ao usuário a tela inicial do game (Figura 10). Ao clicar no botão “Jogar” será apresentada a subtela de Menu (Figura 10), mostrando as opções de jogo que poderão ser escolhidas. O usuário poderá jogar com o “Alfabeto”, com as “Sílabas”, ou com a construção de “Palavras”. Ao iniciar qualquer fase do jogo será comunicado de forma auditiva ao usuário como ele poderá controlar o Avião e o que ele deve fazer para completar a fase, no caso, recolher toda uma sequência de elementos.

Figura 10. Telas do jogo Alfabeto Voador, à esquerda a tela inicial do jogo, à direita o submenu indicando as fases do game.

Ao jogar com o Alfabeto será mostrado o Avião que deverá ser controlado pelo usuário. Na parte superior da tela é mostrada a sequência dos elementos alfabéticos que devem ser coletados (veja Fig. 11). No lado direito do cenário há uma barra de progresso que evolui parcialmente quando um objeto é coletado na sequência correta, esta barra, chegando ao seu máximo, fará o avião sofrer uma transformação, mudando-o para um novo modelo na cor vermelha.

A fase na qual o usuário interage com as Sílabas possui um submenu (Fig. 12), onde o jogador terá que indicar com quais sequências de sílabas deseja jogar. A lógica de jogo nesta fase é a mesma, a diferença cabe somente à parte superior do cenário, onde são mostrados diferentes conjuntos de sílabas que devem ser coletadas. Na fase de construção de Palavras, o cenário é similar às fases anteriores, diferenciando-se em sua parte superior, onde são apresentadas diferentes palavras acompanhadas de figuras que as representam. A palavra, ao ser completada, indicada auditivamente a sua pronúncia.

Figura 11. Jogo Alfabeto Voador, fase do alfabeto.

Figura 12. Submenu, fase das Sílabas, indicar qual família de sílabas para jogar.

4. Avaliação dos Jogos

O processo de avaliação dos jogos ocorreu mediante a apresentação destes aos professores de uma Escola Pública de Ensino Infantil na cidade de Santarém-PA. Esta escola tem seu ensino voltado para alunos do Pré-2 ao 5º ano do Ensino Fundamental.

Do corpo docente da escola, 05 (cinco) professoras se dispuseram a avaliar os jogos digitais, respondendo questionários para avaliação das ferramentas. Ao todo foram 12 questões que, dependendo do questionamento, tinham quatro níveis de indicação, indo desde a baixa aceitação ao quesito avaliado (nenhuma) até a sua aceitação por completo excelente.

Conforme os questionários, os professores possuem pouca ou nenhuma dificuldade em lidar com as disciplinas escolares presentes nos jogos. De modo similar, quando questionados sobre a dificuldade de seus alunos em relação aos conceitos mostrados nos jogos, indicaram que alguns têm dificuldade em lidar com os assuntos. Os docentes quando questionados se gostaram dos jogos e se os consideravam divertidos, afirmaram que são excelentes.

Segundo os professores, ambos os jogos têm capacidade para prender a atenção dos alunos, fornecem conteúdo que estimula o foco na atividade e também estão relacionados com a aprendizagem da disciplina com qual trabalham. Assim, segundo o processo de avaliação e através da apresentação dos games ao público de interesse, verifica-se que os jogos digitais desenvolvidos a partir deste trabalho são adequados para serem usados no processo de ensino e aprendizagem das disciplinas a que se destinam.

5. Considerações Finais

Ao longo deste trabalho discutiu-se a utilização de jogos digitais como ferramentas de complementação ao processo de ensino-aprendizagem e, neste sentido, foi realizado o desenvolvimento de dois jogos digitais que podem ser utilizados como ferramentas auxiliares na fixação de conteúdos de disciplinas do currículo escolar. No decorrer da apresentação dos conceitos, quis-se destacar que os jogos digitais são geralmente vinculados à diversão, mas que o prazer proporcionado por eles pode ser transferido para o aprendizado.

Como trabalhos futuros, os jogos aqui expostos podem ainda receber versões que os possibilitem serem executados em dispositivos móveis e via internet. Pode-se ainda investir na adição de novos componentes aos jogos, como a apresentação de novos desafios, avatares, efeitos visuais e sonoros.

Referências

Barros, L. D. (2012). *Análise de um jogo como recurso didático para o ensino da geometria: jogo dos polígonos.* Recife, PB.

Lima, A. S. (2014). *Proposta de objeto de aprendizagem para o ensino das formas geométricas.* XI Simpósio de Excelência em Gestão e Tecnologia.

Menezes, A. L., Muzatti, L. A., & Taquaritinga, F. d. (2016). JOGOS NO ENSINO DA MATEMÁTICA: o uso de aplicativos como estratégia de aprendizagem . São Paulo: Faculdade de Tecnologia de Taquaritinga.

Pizarro, E. M. (2012). *Jogo Digital: um auxílio no processo de alfabetização.* Porto Alegre, Brasil: CINTED/UFRGS.

Potapczuk, D. d. (2013). K- engine: desenvolvimento do motor do jogo-simulador kimera cidades imaginárias. Salvador, BA: UNEB/GESTEC.

Silva, F. d., Costa, F. P., & Santos, C. L. (2008). *Concepção e realização de um jogo educativo no contexto da aprendizagem colaborativa.* SBC - Proceedings of SBCGames 08: Game e Culture Track.

Vargas, J. C. (2016). AlfaBrinque: Desenvolvimento de jogo digital móvel enquanto ferramenta de auxílio para a alfabetização. Corumbá, MS: UFMS.

3. Combining Naive Bayes

Chapter 3

Combining Naive Bayes and Latent Information For Word Prediction

Henrique X. Goulart, Mauro D. L. Tosi, Daniel Soares-Gonçalves, Paulo Sergio Rodrigues and Guilherme Wachs-Lopes

Abstract

The Natural Language Processing area has been given too much attention by researchers. One of the main motivation beyond this interest is related to the word prediction problem, which states that given a set of words in a sentence, one can recommend the next word. In literature, this problem is solved by methods based on the syntactic or semantic analysis. The models, that considers both methods, can demand high computational effort, which can make the model infeasible for certain applications. With the advance of the technology and mathematical models, it is possible to develop faster systems. This work proposes a hybrid word suggestion model, based on Naive Bayes and Latent Semantic Analysis, considering neighboring words around unfilled gaps. Results show that this model achieves 44.2% of accuracy in the MSR Sentence Completion Challenge.

1.1. Introduction

Word prediction is a word processing feature that aims reduce the number of keystrokes necessary for typing words [Aliprandi et al. 2008]. Usually, these models predict the next word given a set of words based on a context. Because of that, the Natural Language Processing (NLP) area, which performs tasks such word prediction through understanding and interpretation of texts and speeches, became popular.

One of the first NLP approaches in the word prediction area is the Naive Bayes (NB), an method that assumes the conditional independence of its variables. It requires less memory and processing than approaches that consider this dependence. However, the Naive Bayes accuracy tends to be worst than those because of its knowledge loss during the independence assumption [Russell and Norvig 2009].

The NLP area still has researches for an accurate method that could be implemented and predict in an applicable time. In this context, the latent semantic analysis

(LSA) was created. This technique is used to semantically analyze texts through the relationship between the words in different text levels, as phrases, paragraphs, among others [Landauer, Foltz, and Laham 1998]. Besides the high accuracy of the LSA technique, its inferences still are based only on the text frequency, which means that it does not consider word orders and consequently the text syntax.

With the advances of technology, in terms of memory space and processing time, other methods could be developed and consequently implemented, such as the Deep Learning [Mikolov et al. 2013]. The Deep Learning process achieves a great precision in sentences completion challenges, however, it has some issues. The computational cost and time to train the amount of data needed is higher compared to other approaches. Beside this, since the learned knowledge is in the weights of neurons connections, it is not possible to interpret what was learned, which means that this is a black box model.

There are many word prediction models in the NLP area, however, as far as we know, none of them can accomplish the task of predicting a word with precision in an applicable time, as a human being would do. Therefore, the area still demands development and a hybrid model might help researches to explore new approaches.

This paper proposes a new hybrid model to predict words using two well-known models: the Naive Bayes and the LSA. In addition to this, we optimize parameters used to improve the prediction precision through the Gradient Descent technique.

This paper is organized as following. In the next section, we present the relevant areas backgrounds. Then, in the third section, the methodology used in the model and its optimization is explained. Next, in the fourth section, the experiments performed to qualify the results of the model precision are presented. Then, in the fifth section, the benefits and issues are discussed in the paper conclusion.

1.2. Background

Table 1.1. Related models.

Work	Naive Bayes	Other n-gram	LSA	Other model	Year
[Hunnicutt and Carlberger 2001]	X				2001
[Al-Mubaid 2003]	X			X	2003
[Kozima and Ito 2004]				X	2004
[Al-Mubaid 2007]	X			X	2007
[Aliprandi et al. 2008]				X	2008
[Zweig, Platt, et al. 2012]		X	X		2012
[Koutný 2012]				X	2012
[Mikolov et al. 2013]				X	2013
[Kleinman, Runnqvist, and Ferreira 2015]				X	2015
[Spiccia, Augello, Pilato, and Vassallo 2015]			X		2015
[Spiccia, Augello, and Pilato 2015]		X	X		2015
[Luke and Christianson 2016]			X		2016
[Cavalieri et al. 2016]	X				2016

In the last decades, the models to predict a word has been improved due to improvements in mathematical models and a grow of computational power. There are many applications and goals for those predictors, e.g. assist text production or reduce the number of keystrokes as stated before.

In NLP area, many models were proposed to predict a word, a list of some related models is presented in Table 1.1. It shows that Naive Bayes and LSA was not used together to predict words.

1.2.1. Naive Bayes

The Naive Bayes is a probabilistic model used in Natural Language Processing (*NLP*) as a n-gram, developed through Bayesian networks that are vastly used in the machine learning area, as stated in [Russell and Norvig 2009].

This model states that its variables can be divided in to *cause* and *effect* behaviors; thereby, it can be assumed that the *effects* are conditionally independent between themselves, which reduces the computational cost of the model. This model is mathematically represented in Equation (1), in which, the effect and the cause are represented by e and c_i respectively. Thus, using this model it is not necessary to build a joint probability table.

$$P(e|c_1,\cdots,c_n) = \frac{P(e)\prod_{i=1}^{n} P(c_i|e)}{\gamma} \tag{1}$$

1.2.2. Latent Semantic Analysis

The latent semantic analysis (LSA) has been vastly used in systems that analyze textual contents [Zupanc and Bosnić 2017], since it performs a comparison between words to infer.

The LSA can use any cohesive textual level, such as phrases, paragraphs, entire documents in its training set [Coccaro and Jurafsky 1998]. The information (text) of this textual levels contains the semantic relationship between the words in it. To store these relations, a table is constructed (*relationship table*) containing the frequency ($f_{i,j}$) that any word i appeared in a textual level j as illustrated in Table 1.2.

Table 1.2. Example of a relationship table with relations between words and paragraphs.

	Paragraph 1	**Paragraph 2**	**Paragraph 3**	**Paragraph 4**
Word 1	2	7	0	7
Word 2	1	2	1	1
Word 3	0	3	4	5
Word 4	2	7	0	5

With the *relationship table*, comparisons between words can be performed to discover similar words. Therefore, it is necessary to compute the distance between the *relationship table* lines using some metrics [Zhila et al. 2013].

A common problem of the LSA is the dimension of the *relationship table*, that is usually sparse and consumes unnecessary memory space. A solution used in the area for this issue [Spiccia, Augello, Pilato, and Vassallo 2015][Coccaro and Jurafsky 1998] is the usage of the Singular Value Decomposition (SVD) technique. With it, it is possible to reduce the *relationship table* size without losing all the original information, keeping it dense. In this work, the reduced table is noted as *Semantic Reduced Table* (*SRT*).

1.2.3. Gradient Descent

The gradient descent is a mathematical optimization algorithm used to modify variables values based on their contribution to minimize an error function [Witten et al. 2016]. Often, this technique is used in prediction models such as: neural networks, deep learning, bayesian learning, among many others.

To perform the optimization, the error function has to be differentiable, since the gradient descent algorithm uses derivatives. The most used error function is the squared-error [Witten et al. 2016], represented by Equation (2), in which $f(x)$ is the model prediction, y is the ideal value of the prediction (usually 0 or 1) and $\frac{1}{2}$ is used to simplify the further derivatives calculus, as it will be dropped in the next processes.

$$E = \frac{1}{2}(y - f(x))^2 \tag{2}$$

The variables values changes are proportional to the error function gradient [Homod et al. 2012]. Thereby, to minimize the error, the variables have to be updated proportionally in the opposite direction of gradient function. To smooth the optimization updates, a factor η is often used to control and keep the current knowledge of the model, compared with the recent information acquired.

1.3. Proposed Model

This paper proposes a hybrid word prediction model that performs inferences based on Naive Bayes and Latent Semantic Analysis (LSA) theories. The methodology used to develop the proposed model is divided into three stages: Training, Optimization and Inferences, illustrated in Figure 1.1. Those steps are described and analyzed in this section.

1.3.1. Training

In order to train this hybrid model, a textual database is required. Therefore, the Project Gutenberg database [*Project Gutenberg* 2017] was used to accomplish this pre-requisite as it is vastly used in literature such as in [Spiccia, Augello, Pilato, and Vassallo 2015][Gubbins and Vlachos 2013][Zweig, Platt, et al. 2012]. This database is a set composed by 522 19th Century literature books was used in this paper.

1.3.1.1. Naive Bayes Graph

The co-occurrences patterns which represents the Naive Bayes network can be stored in a set of graphs $G = (g_0, g_1, ..., g_{d-1})$, where nodes represent words and edges represent the number of times that each pair of words co-occurred in a same textual level. Each graph g_i with $0 <= i <= n-1$ how many words should be skipped between two co-occurrences.

For instance, consider the phrase (w_1, w_2, w_3, w_4) and the graph g_1. In this case, the first connection will be represented by w_1 and w_3, since $i = 1$ and one word have to be skipped. This enables longer interactions representations with a graph model.

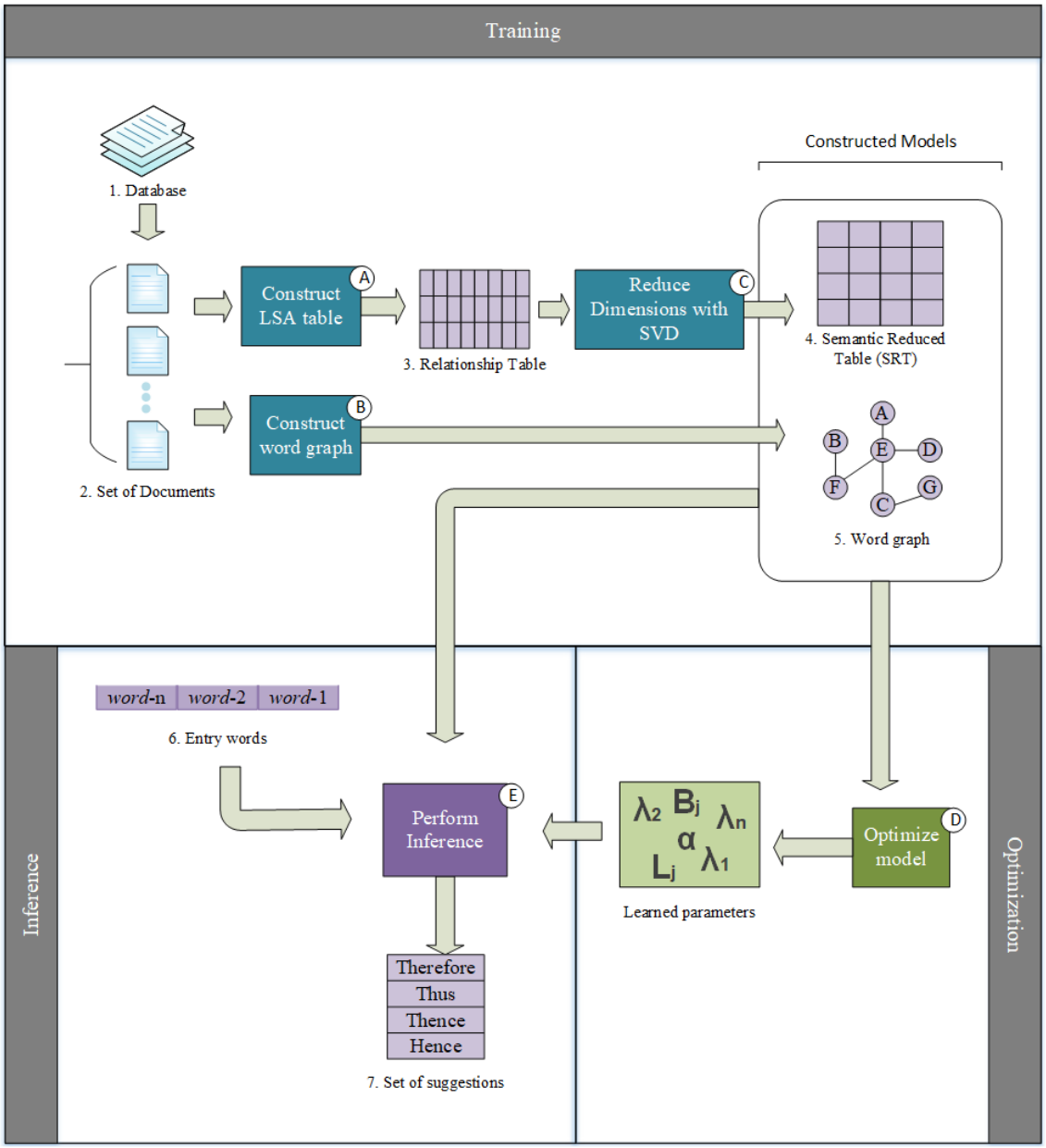

Figure 1.1. Proposed model

1.3.1.2. Latent Semantic Analysis (LSA)

This step consists of building the *relationship table*. Firstly, it is necessary to adopt the textual level to be used in the LSA. In this paper, only sentences with more than 4 nonstop-words are used. This measure was established to gather only the most semantic relevant phrases[1].

To construct the *relationship table*, each different word in the database is repre-

[1]The source-code used as part of implementation of this sub-model can be found in following link: https://github.com/chiawen/sentence-completion

sented as a row and each textual level as a column. The value of each table cell $c_{word,t}$ is defined by the number of times that the word is present in the textual level t.

1.3.2. Inferences

Based on the n previously used words in a text, inferences can be performed in the constructed models to predict a word usage probability in the analyzed context. This section will describe these inferences using the notation $prev_{0:n-1}$ to represent all the n previous words.

1.3.2.1. Naive Bayes inference

The bayesian inferences are computed to establish the $P(sugg_j|prev_0,...,prev_{n-1})$, in which $sugg_j$ is a possible word recommendation (suggestion) from $prev_0,...,prev_{n-1}$ history. This value is then normalized by all the inferences possibilities and is represented by B_j.

In order to boost the Naive Bayes inference, variables responsible to weight each partial probability were inserted in the model, noted as λ_i, in which i refers to the distance between the weighted and the analyzed words.

In Equation 3, B_j represents the weighted model, in which $P^0(prev_0|sugg_j)$ is inversely weighted by the $\lambda_{1,...,n-1}$. Thereby, the value of $\lambda_{1,...,n-1}$ are normalized between themselves, and the equation energy is preserved.

$$B_j = \frac{P(sugg_j)P^0(prev_0|sugg_j)^{\frac{1}{\lambda_1 * ... * \lambda_{n-1}}} \prod_{i=1}^{n-1} P^i(prev_i|sugg_j)^{\lambda_i}}{\gamma} \tag{3}$$

1.3.2.2. Latent Information Inference

The LSA inference L_j is computed using the similarity of a candidate word j (present in SRT) to all $prev$ words as illustrates the Equation (4), the inverse of the distance vector values are summed, as the distance is inversely proportional to the semantic similarity of the words; then, it is applied a normalization based in n.

$$L_j = \frac{1}{n} \cdot \sum_{i=0}^{n-1} \frac{1}{\|j-i\|_2 + 1} \tag{4}$$

1.3.2.3. Hybrid Inference

The trained networks achieved through the Naive Bayes and LSA models, and their respective inferences, Equations (3) and (4), output different pattern results, which affects the hybrid model development. The probability variance of the Naive Bayes inferences is much larger than the LSA ones.

In order to successfully merge the models, it is necessary to establish an output pattern to equalize them. Therefore, the probability values of the next word to be inferred, from both models, are sorted in ascending order crescent in two vectors. Thereby, those probabilities are replaced by their vector index and normalized by the sum of all indexes, as illustrated below.

Thus, the Naive Bayes and LSA inferences are weighted using a parameter $alpha$, which creates the hybrid model, that is represented in Equation (5).

$$\theta_j = \alpha B_j + (1-\alpha)L_j \tag{5}$$

1.3.3. Optimization

Once Naive Bayes and LSA networks are trained, the inferences can be performed. However, those models provide a different set of probabilities (B_j and L_j). Considering the hybrid model of Equation (5), there is a set parameters that can be optimized in order to provide an ideal weight between the models, improving the recommendation precision. Moreover, the values of λ parameters can also be optimized, achieving a better weight between the partial probabilities of Naive Bayes.

Therefore, optimization through the gradient descent technique was performed. The set of optimized parameters was composed by α, that represents the relevant percentage of the Naive Bayes compared to the LSA; and $\lambda_1, ..., \lambda_{n-1}$, that express the weight/relevance relation between the words distances in the Naive Bayes inference.

1.4. Results

In order to validate the proposed model, it was applied the experiment MSR Sentence Completion Challenge [Zweig and Burges 2011]. This experiment consists of the 1040 sentences with a missing word. Moreover, it is provided 5 options to choose a word for complete the sentence.

In the MSR Sentence Completion Challenge, it is possible that the missing word is in the middle of sentence. For this case, the our bayesian model also considers the posterior words from the gap.

To guarantee cross-validation on experiments, the MSR Sentence Completion Challenge was separated into 5 groups (5-fold cross-validation). Therefore, one of these groups was used for test and the others for optimization. Furthermore, is possible to have 5 different and independent configurations for experiments.

1.4.1. Error Variation

In order to observe the proposed optimization using a 3-gram history, it was created a scenario where the parameters λ and α were initialized randomly. Furthermore, the optimization groups were disposed in a cyclic loop (i.e. epochs), thus extending the optimization.

The Figure 1.2 shows the error variation in each configuration. It is possible to observe that the error is minimized for all configurations. This error metric is obtained through the sum of all individual sentences errors.

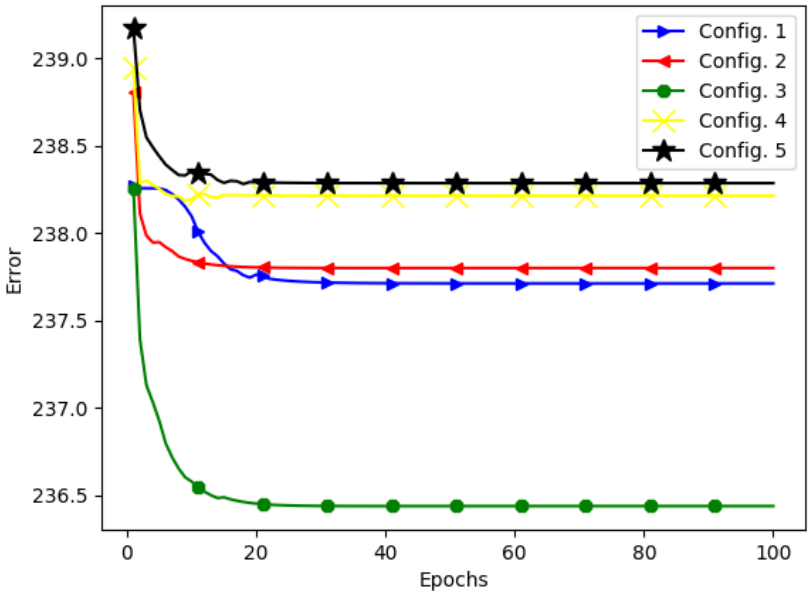

Figure 1.2. Variation of error throughout the epochs.

1.4.2. Tests and Comparative

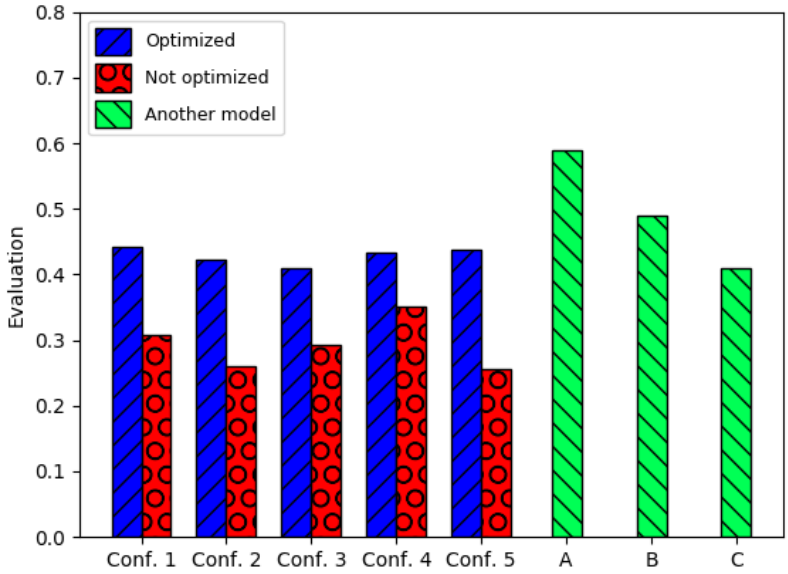

Figure 1.3. Results of MSR Sentence Completion Challenge and comparative with (A) RNNLM + Skip-gram [Mikolov et al. 2013], (B) LSA [Zweig and Burges 2011] and (C) N-gram [Gubbins and Vlachos 2013].

As illustrated in Figure 1.3, the proposed model in this work using a 3-gram history is superior to another n-gram model which considers history between 5-gram; but it is worse than LSA model, which considers all words in the sentence for inference; also it is worse than neural network model. However, the proposed model provides recommendations considering syntactic and semantic features using less training data than a neural network model.

1.5. Conclusion

In this work, it was proposed a hybrid model to complete sentences employing Naive Bayes and LSA models. This hybrid model uses a set words to infer the next word considering each sub-model individually.

The gradient descent was used to optimize the parameters of the model. During the optimization, the error function was optimized with $\alpha = \{0.2 \sim 0.4\}$. This means that LSA had more influence over the final recommendation.

During the tests, the proposed model showed promising results compared to other state-of-art related models. Even not providing better results than actual state-of-the-art, the proposed model brought relevant results, considering syntactic/semantic rules among words and consuming a fraction of training required by a model based on Deep Learning.

Thereby, the proposed model in this work achieved the objectives providing a great time $\times$ accuracy relation which can be improved, thus creating new perspectives in NLP as also different applications such: a better word prediction model for mobile platforms, improve the conversation ability of robots, enhance search engines and improve the social interaction of impaired people.

References

Aliprandi, Carlo et al. (2008). *Advances in NLP applied to Word Prediction*.

Cavalieri, Daniel C et al. (2016). "Combination of language models for word prediction: an exponential approach". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24.9, pp. 1481–1494.

Coccaro, Noah and Daniel Jurafsky (1998). "Towards better integration of semantic predictors in statistical language modeling." In: *ICSLP*, pp. 2403–2406.

Gubbins, Joseph and Andreas Vlachos (2013). "Dependency Language Models for Sentence Completion." In: *EMNLP*. Vol. 13, pp. 1405–1410.

Homod, Raad Z et al. (2012). "Gradient auto-tuned Takagi–Sugeno Fuzzy Forward control of a HVAC system using predicted mean vote index". In: *Energy and Buildings* 49, pp. 254–267.

Hunnicutt, Sheri and Johan Carlberger (2001). "Improving word prediction using markov models and heuristic methods". In: *Augmentative and Alternative Communication* 17.4, pp. 255–264.

Kleinman, Daniel, Elin Runnqvist, and Victor S Ferreira (2015). "Single-word predictions of upcoming language during comprehension: Evidence from the cumulative semantic interference task". In: *Cognitive psychology* 79, pp. 68–101.

Koutný, Michal (2012). "Word prediction using language models". In:

Kozima, Hideki and Akira Ito (2004). "A scene-based model of word prediction". In: *Available at citeseer. ist. psu. edu/97151. html*.

Landauer, Thomas K, Peter W Foltz, and Darrell Laham (1998). "An introduction to latent semantic analysis". In: *Discourse processes* 25.2-3, pp. 259–284.

Luke, Steven G and Kiel Christianson (2016). "Limits on lexical prediction during reading". In: *Cognitive Psychology* 88, pp. 22–60.

Mikolov, Tomas et al. (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Al-Mubaid, Hisham (2003). "Context-based word prediction and classification". In: *Proceedings of the 18th International Conference on Computers and their Applications CATA'2003*, pp. 384–388.

— (2007). "A Learning-Classification Based Approach for Word Prediction". In: *Int. Arab J. Inf. Technol.* 4, pp. 264–271.

Project Gutenberg (Nov. 2017). URL: www.gutenberg.org.

Russell, Stuart and Peter Norvig (2009). *Artificial Intelligence: A Modern Approach*. 3rd. Upper Saddle River, NJ, USA: Prentice Hall Press. ISBN: 0136042597, 9780136042594.

Spiccia, Carmelo, Agnese Augello, and Giovanni Pilato (2015). "A Word Prediction Methodology Based on Posgrams". In: *International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management*. Springer, pp. 139–154.

Spiccia, Carmelo, Agnese Augello, Giovanni Pilato, and Giorgio Vassallo (2015). "A word prediction methodology for automatic sentence completion". In: *Semantic Computing (ICSC), 2015 IEEE International Conference on*. IEEE, pp. 240–243.

Witten, Ian H et al. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Zhila, Alisa et al. (2013). "Combining Heterogeneous Models for Measuring Relational Similarity." In: *HLT-NAACL*, pp. 1000–1009.

Zupanc, Kaja and Zoran Bosnić (2017). "Automated essay evaluation with semantic analysis". In: *Knowledge-Based Systems*.

Zweig, Geoffrey and Christopher JC Burges (2011). *The Microsoft Research sentence completion challenge*. Tech. rep. Technical Report MSR-TR-2011-129, Microsoft.

Zweig, Geoffrey, John C Platt, et al. (2012). "Computational approaches to sentence completion". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, pp. 601–610.

4. Cluster Specialization

Chapter 4

Exploiting Cluster Specialization into Linear Weighted Hybrid Recommender Systems

Miller Horvath, Matheus Henrique C. Zamberlan, Vitor Finati, Guilherme A. Wachs-Lopes and Paulo S. Rodrigues

Abstract

Recommender Systems effectively filter huge amounts of data and deliver personalized content recommendations. Clustering methods are commonly adopted by Recommender Systems aiming at reducing computational cost. On the other hand, this work raises an hypothesis that, if the clustering method provides high data discrimination, exploiting clusters may lead to a substantial accuracy improvement. An initial test of this hypothesis is provided adopting K-means and Collaborative Filtering as the clustering and recommender algorithms, respectively. Although the experimental results showed a slight accuracy improvement, compared with the traditional Linear Weighted Hybrid approach, this work contribution gives a new perspective of exploring clustering methods to accuracy purposes. A future work will adopt more robust clustering and recommender algorithms.

1.1. Introduction

Currently, the amount of content available on Internet is vast and it is continually increasing. Many on-line services provide an unmanageable list of content. As a result, users usually struggle on searching the content of their interest, which might cause frustration and even lead them to stop using those services. Thus, the well-known Recommender Systems, originated from the work in [Goldberg et al., 1992], have emerged aiming at tackling this issue. Collaborative Filtering basically analyses the items (content available in a given service, such as movies, songs, videos and news) that users have consumed in the past, perceiving the items they have enjoyed and other users who have similar taste, based on similarity measures, in order to predict the items they may enjoy in the future [Ricci et al., 2011].

After the Netflix Prize [Bennett et al., 2007], both the academic and commercial communities increased their interest in developing Recommender Systems. It can be

noted its importance to commercial purposes, since great companies, as Netflix[1], Amazon[2], YouTube[3], and Spotify[4], make use of those systems to recommend relevant content to their users.

Combining different types of recommender algorithms into hybrid recommenders is known to provide performance improvements [Burke, 2002]. Linear Weighted Hybrid Recommenders (LWHR) [Burke, 2002] combines two or more recommender algorithms by computing a weight score to each algorithm that increases the recommendation performance. On the other hand, clustering algorithms are applied into Recommender Systems in order to reduce the amount of data that must be matched in the similarity computation step [Amatriain and Pujol, 2015].

In this work, it is raised the hypothesis that clustering algorithms may lead to accuracy improvement by specializing LWHR into each data cluster. Thus, rather than computing the weight scores that maximizes the accuracy in the entire data, the cluster specialization computes the weight score that maximizes the accuracy in each cluster individually. Let us suppose that the traditional LWHR scores are assigned to each cluster in the dataset. It is reasonable to consider that, if the scores of a given cluster is changed, achieving a local accuracy improvement, a global accuracy improvement would also be achieved. However, in order to achieve a substantial performance improvement, the clustering algorithm must highly discriminate the dataset. An initial evaluation of our hypothesis is presented, using K-Means algorithm [Lloyd, 1982] for clustering and two derivations of Recommender Systems algorithms, which are User-Based [Schafer et al., 2007] and Item-Based Collaborative Filtering [Sarwar et al., 2001].

The remainder of this work is organized as follows. First, some of the academic contributions in Recommender Systems that are related to this work are presented in Section 1.2. Second, the methodology is presented in Section 1.3. Third, in Section 1.4, the experimental evaluation procedure is formalized and the obtained results are presented and discussed. Finally, in Section 1.5, the work is concluded and future works are presented.

1.2. Related Work

Recently, it has been common to see Recommender Systems using clustering methods to reduce computational cost of processing huge amounts of data. In [Zhou et al., 2016], the Markov Clustering Algorithm is used in order to group data, based in keywords, and find patterns to recommend on-line videos. It is more interesting when its verified that each cluster usually became restricted to a specific theme. The work in [Ghenname et al., 2013] uses keywords and semantic analysis to group e-learning content and recommend it to a social network users. For the purpose of testing the efficiency of the method proposed, the clustered separation was compared with human separation and presented a success rate of more than 89%. To improve the route of taxi drivers, probabilistic techniques, Filtering and Bisected Clustering Algorithm were used in [Yuan et al., 2013] to create

[1] https://www.netflix.com
[2] https://www.amazon.com/
[3] https://www.youtube.com
[4] https://www.spotify.com

a Recommender System to increase taxi drivers profit, reducing their idle time. Tests performed in Beijing, China, showed better profit and efficiency to taxi drivers services. In [Gupta and Gadge, 2015], the K-means algorithm is used to separate users based in their demographic data. Comparing with other models that also use Collaborative Filtering, the proposed algorithm demonstrated greater coverage and lower Mean Absolute Error (MAE). In addition, it deals with the cold-start problem, which is a problem caused by lack of information of either users or items that are new in the system, faced by many Recommender Systems. In [Gupta and Gadge, 2014], a combination of Item-Based Collaborative Filtering and Demographics-Based User Clusters is used to create a Hybrid Recommender System that considers demographic information to deal with traditional cold-start problems. It uses K-means to cluster users and process all data off-line to improve scalability.

In order to improve the recommendation quality and precision, Hybrid Recommender Systems have been increasingly used. A model based in Fuzzy Clustering and Collaborative Filtering was proposed in [Verma et al., 2013]. This technique showed more efficiency when compared with others that used only K-means clustering. Hybrid Recommender Systems are also used in [Tejeda-Lorente et al., 2014] to recommend relevant research sources in digital university libraries. The work in [Bostandjiev et al., 2012] proposes a Hybrid Recommender System using a combination of different sources of data and user interaction, using Content-Based Filtering and Collaborative Filtering. In [Zanotti et al., 2016], accuracy improvements are achieved combining Memory-Based Collaborative Filtering and Neural Network Model-Based Collaborative Filtering models into a Linear Weighted Hybrid Recommender.

1.3. Methodology

Collaborative Filtering (CF) algorithms analyze the users behavior in the system over time. This behavior is represented by a rating matrix, which contains the ratings users have assigned to items they have consumed. Those ratings can be binary (like or dislike) or discrete (a score between 1 and 5). Therefore, it is possible to represent users and items characteristics as n-dimensional vectors. Using these vectors to compute user-to-user and item-to-item similarities, through similarity measures, and use the most similar elements in order to predict the rating users would assign to items they have not consumed yet.

Firstly, the traditional User-Based Collaborative Filtering (UBCF) and Item-Based Collaborative Filtering (IBCF) models are adopted. By applying K-Means clustering algorithm in the rating matrix, clustering both users and items, two more models are obtained, which are the User-Clustered Collaborative Filtering (UCCF) and Item-Clustered Collaborative Filtering (ICCF). The difference between the traditional and the cluster-based models lies in the similarity computing step. In the traditional models, the k-most similar elements are computed considering the entire dataset. On the other hand, the cluster-based models narrow the similarity computation between elements inside the same cluster only.

In the next step, the k-Nearest Neighbors (k-NN) algorithm [Altman, 1992] is used to compute the list of k-most similar elements in the dataset, which is called *similarity list*. The similarity measure adopted is cosine similarity. The cosine similarity between two

n-dimensional vectors A and B ($sim(A,B)$) is given by:

$$sim(A,B) = \frac{A \cdot B}{\|A\| \, \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

Being A_i and B_i the i-th rating of A and B respectively.

Then, based on the similarity list, the ratings that users would assign to items they have not consumed yet are predicted. To predict the rating a user u would assign to an item i, the User-Based and Item-Based rating predictions are adopted, given by the Equations 2 and 3 respectively. In these equations, U stands for the list of users similar to user u, I stands for the list of items similar to item i, and the function $rat(a,b)$ stands for the observed rating a user a have assigned to an item b.

$$ub\_pred(u,i) = \frac{\sum_{u' \in U} sim(u,u') rat(u',i)}{\sum_{u' \in U} sim(u,u')} \tag{2}$$

$$ib\_pred(u,i) = \frac{\sum_{i' \in I} sim(i,i') rat(u,i')}{\sum_{i' \in I} sim(i,i')} \tag{3}$$

Next, the weight score of each model is computed, maximizing the accuracy for the entire dataset. The accuracy is measured by the Root Mean Squared Error (RMSE) between the predicted and observed ratings, given by Equation 4. The lower the obtained error, the higher the accuracy. The Coverage in each model is used to evaluate the percentage of ratings the Recommender Systems were able to predict.

$$RMSE = \sqrt{\frac{1}{|N|} \sum_{n \in N} (\, rat_n - rat'_n)^2} \tag{4}$$

In Equation 4, N stands for a set of ratings pairs, $|N|$ stands for the number of rating pairs in N, and rat_n and rat'_n stand for the observed and the predicted ratings respectively.

Finally, based on the dataset clustering of users and items, the data is organized into the so-called *predicting contexts*, which is the foundation of the proposed cluster specialization of LWHR. Let us say that C_u is the set of users clusters, composed by N clusters, and C_i is the set of items clusters, composed by M clusters, this means there are $N \times M$ predicting contexts, see Fig. 1.1. For all user clusters $uc \in C_u$ and all item clusters $ic \in C_i$, the predicting context that relates uc and ic is composed by the rating predictions users in uc would assign to items in ic. Thus, the cluster specialization of LWHR computes the set of weight scores that maximizes the accuracy of recommendations for each predicting context, resulting in $N \times M$ sets of weight scores.

1.4. Experiments, Results and Discussion

The experiments were developed in Python programming language. The Python's scikit-learn library [Pedregosa et al., 2011] implementation of both k-NN and k-Means algorithms were adopted.

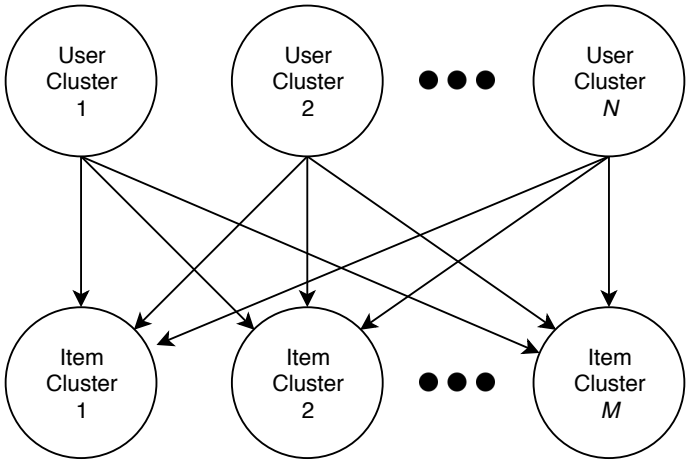

Figure 1.1. Illustrates the predicting contexts. The arrows stands for each possible predicting contexts.

1.4.1. Dataset

In order to evaluate our proposal, a widely adopted dataset for Recommender Systems was used, called MovieLens 1M Dataset [Harper and Konstan, 2016]. Built in a movie domain, this dataset has about 1 million ratings, 6000 users and 4000 movies. Each user in the dataset rated at least 20 movies. Ratings are made on a 5-star scale (from 1 to 5 with whole-star ratings only).

1.4.2. Traditional Collaborative Filtering

The experiments were started by analyzing the impact of varying the number of neighbors k adopted into the tradition CF, considering $k = \{5, 10, 20, 50, 100, 200\}$. For both UBCF and IBCF, the highest coverage and the lowest RMSE were achieved with $k \geq 50$. Thus, only these values of k were used in the following experimental steps.

1.4.3. Cluster-Based Collaborative Filtering

Since k-Means algorithm, which requires to preset the number of desired clusters, was adopted, we also explored the impact of varying the number of clusters c, considering $c = \{5, 10, 50, 100, 250\}$. Since both UCCF and ICCF achieved coverage close to 100%, see Fig. 1.2, and the lowest RMSE with five users and items clusters, only $c = 5$ was kept in the remainder experiments.

1.4.4. Comparing Recommender Models

After exploring all the four models individually, the Traditional and Cluster-Specialized Linear Weighted Hybrid Recommenders were built. The results are presented in Fig. 1.3. In the adopted dataset, the items clustering provides a slight increase in predicting accuracy, for higher values of k ($k \geq 50$), when compared to the traditional IBCF. On the other hand, the clustering of users presented a drop in accuracy when compared to traditional UBCF. As we expected, the hybrid models showed a considerable accuracy improvement. In conclusion, the cluster specialization proposed (CS-LWHR) offered just a slight performance improvement when comparing with the traditional LWHR algorithm, suggesting

that the clustering applied did not discriminate the users and items in the dataset in a way that the cluster specialization would provide substantial improvements on recommender accuracy.

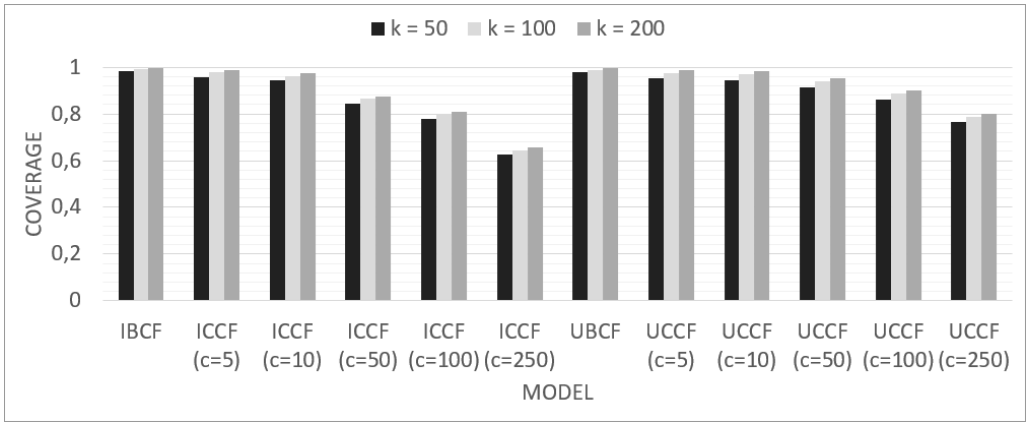

Figure 1.2. Presents the coverage of the following models: Item-Based Collaborative Filtering (IBCF), Item-Clustered Collaborative Filtering (ICCF), User-Based Collaborative Filtering (UBCF), User-Clustered Collaborative Filtering (UCCF).

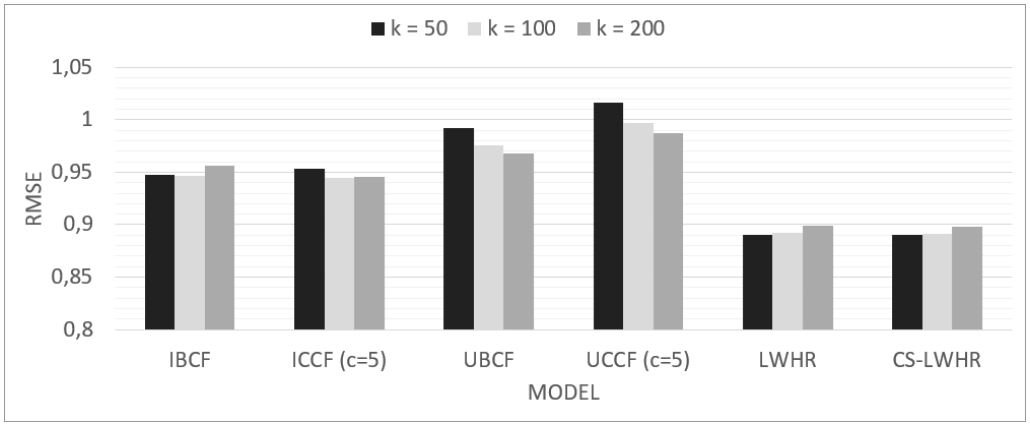

Figure 1.3. Presents the Root Mean Squared Error (RMSE) of the following models: Item-Based Collaborative Filtering (IBCF), Item-Clustered Collaborative Filtering (ICCF), User-Based Collaborative Filtering (UBCF), User-Clustered Collaborative Filtering (UCCF), Linear Weighted Hybrid Recommender (LWHR), and Cluster-Specialized Linear Weighted Hybrid Recommender (CS-LWHR).

1.5. Conclusion and Future Work

Recommender Systems techniques filter huge amounts of data, discarding undesirable content and delivering personalized recommendations to users. Its potential of being applied at different domains resulted in a widely adoption of this systems for commercial purposes. Furthermore, the academic community has shown interest in Recommender Systems as well in the past few years. Many works explored Hybrid Recommender Systems to improve accuracy and clustering techniques to reduce computational cost of Recommender Systems.

In this work, we explored clustering techniques with a different goal, which is increase the accuracy of Hybrid Recommender Systems. This work proposal relies on the idea that, if the elements assigned to the same cluster are very alike, computing the weight scores that maximizes the performance of the system to each cluster would substantially outperform the weight scores that maximizes the non-clustered dataset. Therefore, the proposed hypothesis suggests that the higher is the data discrimination achieved by the clustering method, the better is the accuracy improvement provided by the cluster specialization of LWHR, compared to the traditional LWHR.

We adopted classic techniques of clustering and recommender systems, which are k-Means and Collaborative Filtering respectively. Even though the proposed CS-LWHR has not achieved expressive improvements compared to traditional LWHR, our work still presents a valuable contribution since it proposes a new perspective of applying clustering for accuracy improvement rather than for reducing computational cost only. Moreover, the results suggests a lack of data discrimination provided by the clustering method in this dataset. In a future work, more algorithms for both clustering and recommendation, such as Self-Organizing Maps and Model-Based Collaborative Filtering, will be adopted in order to better investigate the proposed CS-LWHR.

References

[Altman, 1992] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.

[Amatriain and Pujol, 2015] Amatriain, X. and Pujol, J. M. (2015). Data mining methods for recommender systems. In *Recommender systems handbook*, pages 227–262. Springer.

[Bennett et al., 2007] Bennett, J., Lanning, S., et al. (2007). The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York, NY, USA.

[Bostandjiev et al., 2012] Bostandjiev, S., O'Donovan, J., and Höllerer, T. (2012). Tasteweights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 35–42. ACM.

[Burke, 2002] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370.

[Ghenname et al., 2013] Ghenname, M., Abik, M., Ajhoun, R., Subercaze, J., Gravier, C., and Laforest, F. (2013). Personalized recommendation based hashtags on e-learning systems. In *ISKO-Maghreb, 2013 3rd International Symposium*, pages 1–8. IEEE.

[Goldberg et al., 1992] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70.

[Gupta and Gadge, 2014] Gupta, J. and Gadge, J. (2014). A framework for a recommendation system based on collaborative filtering and demographics. In *Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on*, pages 300–304. IEEE.

[Gupta and Gadge, 2015] Gupta, J. and Gadge, J. (2015). Performance analysis of recommendation system based on collaborative filtering and demographics. *Communication, Information & Computing Technology (ICCICT), 2015 International Conference*, pages 1–6.

[Harper and Konstan, 2016] Harper, F. M. and Konstan, J. A. (2016). The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):19.

[Lloyd, 1982] Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Ricci et al., 2011] Ricci, F., Rokach, L., and Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer.

[Sarwar et al., 2001] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM.

[Schafer et al., 2007] Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer.

[Tejeda-Lorente et al., 2014] Tejeda-Lorente, Á., Porcel, C., Peis, E., Sanz, R., and Herrera-Viedma, E. (2014). A quality based recommender system to disseminate information in a university digital library. *Information Sciences*, 261:52–69.

[Verma et al., 2013] Verma, S. K., Mittal, N., and Agarwal, B. (2013). Hybrid recommender system based on fuzzy clustering and collaborative filtering. In *2013 4th International Conference on Computer and Communication Technology (ICCCT)*.

[Yuan et al., 2013] Yuan, N. J., Zheng, Y., Zhang, L., and Xie, X. (2013). T-finder: A recommender system for finding passengers and vacant taxis. *IEEE Transactions on Knowledge and Data Engineering*, 25(10):2390–2403.

[Zanotti et al., 2016] Zanotti, G., Horvath, M., Barbosa, L. N., Immedisetty, V. T. K. G., and Gemmell, J. (2016). Infusing collaborative recommenders with distributed representations. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pages 35–42. ACM.

[Zhou et al., 2016] Zhou, R., Khemmarat, S., Gao, L., Wan, J., Zhang, J., Yin, Y., and Yu, J. (2016). Boosting video popularity through keyword suggestion and recommendation systems. *Neurocomputing*, 205:529–541.

5. Reconhecimento de Emoções

Capítulo 5

Reconhecimento de emoções através da voz para integração em uma aplicação web

Eduardo Costa de Paiva e Márcia Aparecida Fernandes

Abstract

Extracting information from facial expressions has been widely used to identify emotions, and with technological advancement other alternatives such as extracting features of voice and writing have also been used to recognize emotions. This article presents the development of a web application capable of classifying emotions through the voice. Two voice databases were used for training (EmoDB and RAVDESS) and a comparison between two classification algorithms was performed: J48 and SMO. The obtained results show that the SMO obtained better performance, for the base RAVDESS the accuracy was of 79.30 % and for base EmoDB the accuracy was of 87.47 %, such result can be improved from the use of other bases. Despite the difficulties there is great viability of the application.

Resumo

A extração de informações das expressões faciais tem sido largamente utilizada para identificar emoções, e com o avanço tecnológico outras alternativas como extração de características da voz e da escrita também foram utilizadas para reconhecer as emoções. Esse artigo apresenta a criação de uma aplicação web capaz de classificar emoções através da voz. Foram utilizadas duas bases de vozes para treinamento (EmoDB e RAVDESS) e foi realizado um comparativo de dois algoritmos de classificação: J48 e SMO. Os resultados obtidos mostram que o SMO obteve melhor desempenho, para a base RAVDESS a acurácia foi de 79.30% e para base EmoDB a acurácia foi de 87.47%, tal resultado pode ser melhorado a partir do uso de outras bases. Apesar das dificuldades há grande viabilidade de uso da aplicação.

1.1. Introdução

A comunicação através da voz fez parte do processo evolutivo da humanidade, logo após as expressões faciais uma das primeiras tentativas de expressar sentimentos e

opiniões foi através da fala. Os seres humanos se comunicam e interagem a todo momento variando suas emoções. Com o avanço frequente de técnicas como Inteligência Artificial e Reconhecimento de Padrões, a interação humano-computador pôde se aproximar das relações entre humanos. Se o computador é capaz de extrair características que identificam a emoção de um usuário, poderá utilizá-las para providenciar uma experiência melhor para o indivíduo.

Para a construção de uma aplicação capaz de reconhecer emoções, se faz necessário saber o que é uma emoção de fato. Existe uma grande pesquisa que especifica detalhadamente a natureza desse estado, chamada de a "Tradiçao Psicológica"(*Psychological Tradition*)[1] concebida por grandes nomes de diferentes áreas como Filosofia (Rene Descartes), Biologia (Charles Darwin) e Psicologia (William James). Uma problemática é que não existe exatamente um conjunto de emoções básicas. O psicólogo Paul Ekman [2] definiu um conjunto de seis emoções universais, dentre elas: Raiva, Nojo, Medo, Felicidade, Tristeza e Surpresa. Todas as demais emoções seriam variações ou uma mistura das demais. Robert Plutchik [3] propôs um aprimoramento à ideia de Ekman, criando a roda das emoções, que contém 8 emoções básicas (Raiva, Medo, Tristeza, Nojo, Surpresa, Curiosidade, Aceitação e Alegria) e as possíveis variações destas.

Como existem diferentes abordagens de quais emoções devem ser consideradas básicas ou não, deve-se fazer escolhas pragmáticas para o tipo de problema em questão. Neste trabalho foram analisadas duas base de dados, uma em alemão e outra em inglês que possuem as mesmas emoções, exceto por duas diferenças: A emoção Tédio que está presente na base alemã porém não está na base em inglês e as emoções Calma e Surpresa presentes na base em inglês mas que não estão presentes na base em alemão. Essa diferença se dá pura e simplesmente pelo fato de serem diferentes pesquisadores confeccionando cada base, dessa forma cada base se restringiu a um certo grupo de emoções.

A fim de executar o reconhecimento de emoções através da voz, deve-se extrair características deste sinal de áudio. Muito ainda se discute quais parâmetros devem ser utilizados, afinal não se sabe tão bem como nosso cérebro consegue discernir as emoções. Deve-se então criar uma aproximação através de modelos matemáticos. Muitos autores fazem uso de parâmetros como intensidade, frequência fundamental, assim como as propriedades espectrais da voz e coeficientes cepstrais. Neste trabalho foi realizado um estudo sobre características relevantes que possam ser extraídas de um sinal de voz, para se construir uma aplicação capaz de reconhecer emoções através desses sinais. O objetivo futuro deste trabalho é aprimorar um sistema virtual de aprendizagem realizando uma junção da análise de emoções da voz, face e escrita pra prover uma experiência melhor para o aluno utilizando tal sistema.

1.2. Fundamentação teórica

Nesta seção serão abordados aspectos teóricos sobre características do áudio, bases de dados e ferramentas utilizadas para o desenvolvimento e experimentos.

1.2.1. Base de Dados

Foram analisadas duas bases de áudio, a base alemã *Berlin Database of Emotional Speech* [4] e a base em inglês *Ryerson Audio-Visual Database of Emotional Speech and Song* [14]

1.2.1.1. Berlin Database of Emotional Speech (EmoDB)

É uma base de áudios em alemão. Tais áudios são gravações de 10 atores (5 homens e 5 mulheres) simulando 10 sentenças (5 curtas e 5 longas) do cotidiano. As emoções simuladas são as 7 emoções básicas: alegria, nojo, tédio, medo, neutra, raiva e tristeza.

1.2.1.2. Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

É uma base de dados em inglês de vozes e músicas compostas por 24 atores (12 homens e 12 mulheres). Essa base oferece áudios para as seguintes emoções: neutra, calma, felicidade, tristeza, raiva, medo, surpresa e nojo.

1.2.2. Extração de características de um sinal de voz

Existem algumas características que são consideradas relevantes para análise de um sinal de voz. Dentre estas estão o *pitch*, intensidade e os coeficientes mel-cepstrais.

A palavra *pitch* não possui uma tradução, mas pode ser entendida como a frequência medida em um sinal de áudio. Em muitos casos o *pitch* pode ser igual a frequência fundamental conhecida por f_0, que é definida como sendo a menor frequência de um sinal de onda. Mas nem sempre essas duas propriedades serão iguais, a frequência é objetiva, um atributo científico que pode ser medido, já o *pitch* é subjetivo e diz respeito a como o ouvido humano é capaz de diferenciar sons graves e agudos. O *pitch* é relevante para a análise da voz, pois pode facilitar a separação das emoções. Um *pitch* alto com muitas variações indica emoções de alta excitação e um *pitch* baixo com poucas variações reflete emoções de baixa excitação. A intensidade também é outro parâmetro importante, dependente de muitos fatores como locutor, idioma, cultura, o tipo da expressão. Como o *pitch*, uma intensidade alta indica emoções de alta excitação. Um estudo realizado por Murray e Arnott [5] descreve algumas emoções e alguns parâmetros estudados extensivamente na área de psicologia, dentre eles *pitch* e intensidade, a Tabela 1.1 adaptada de [1] [6] resume esse estudo.

Tabela 1.1. Emoções e parâmetros da voz (Murray e Arnott, 1993)

	Raiva	**Felicidade**	**Tristeza**	**Medo**	**Nojo**
Pitch	maior média, maior extensão, mudanças abruptas	maior média, maior extensão	menor média, menor extensão	maior média, maior extensão	menor média, menor extensão
Intensidade	mais alta	mais alta	mais baixa	normal	mais baixa
Velocidade da Fala	ligeiramente mais rápida	mais rápida ou mais lenta	mais lenta	mais rápida	mais lenta
Qualidade da voz	ofegante	ofegante, estridente	ressonante	irregular	resmungante

Outra característica muito utilizada no reconhecimento de emoções através da voz

são os coeficientes mel cepstrais (*Mel Frequency Cepstral Coefficients*)[7][8]. A técnica MFCC faz uma análise de características espectrais de tempo curto baseando-se no espectro da voz convertido para uma escala denominada Mel. A escala Mel relaciona a frequência percebida à sua frequência medida real.

Os seres humanos são muito melhores em discernir pequenas mudanças no tom em baixas freqüências do que em altas freqüências. Incorporar essa escala faz com que a análise fique mais próxima com o que os humanos ouvem. Para se obter os coeficientes, primeiramente é aplicado a Transformada discreta de Fourier, após esta etapa tendo o espectro na escala Mel é efetuado o cálculo do logaritmo desse espectro, pois a intensidade percebida de um sinal de áudio é aproximadamente logarítmica. Depois é efetuado a transformada discreta do cosseno e alguns autores defendem a aplicação de derivadas[18] para a obtenção do vetor característico final. O fluxo dos cálculos mais comuns em relação a estes coeficientes são apresentados na Figura 1.1[16][17]

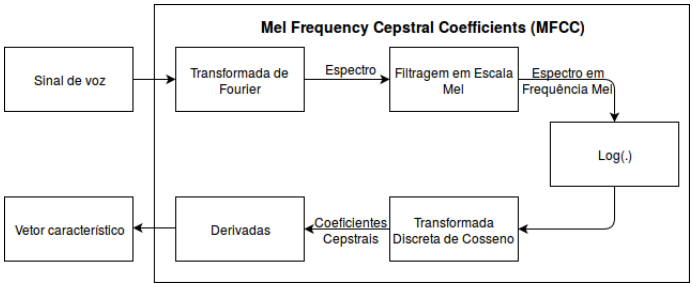

Figura 1.1. Diagrama do algoritmo MFCC [7][19]

1.2.3. Ferramenta para extração de dados da voz - *openSMILE*

The Munich open-Source Media Interpretation by Large feature-space Extraction (*openSMILE*)[9] é uma ferramenta escrita em C++, que tem como principal objetivo a extração de características em processamento de sinais e em aplicações de aprendizado de máquina. O foco primário desta, é nos sinais de áudio. Neste trabalho, a ferramenta *openSMILE* foi utilizada para extrair os parâmetros de um sinal de áudio para classificação de emoções.

1.3. Classificação de um sinal de voz

Nesta seção são descritos os algoritmos para classificação, bem como o método de treinamento utilizado.

1.3.1. Algoritmos e Softwares para classificação

Weka [10] é um *software open-source* com diversos algoritmos de aprendizado de máquina para mineração de dados. Os algoritmos podem ser aplicados diretamente a um conjunto de dados ou chamados através de um código Java. Neste trabalho, o Weka foi utilizado para verificar a partir de um conjunto de algoritmos de classificação, qual possuiria melhor eficácia para reconhecer as emoções através das características extraídas da voz. A Árvore de Decisão J48 foi um dos algoritmos utilizados para os experimentos devido ao fato de já ter sido aplicado em trabalho [11] semelhante a este. O algoritmo J48 é uma implementação *open-source* em Java do algoritmo C4.5 [12] no software Weka.

O SMO (*Sequential Minimal Optimization*)[13] é um algoritmo que foi concebido para ser uma solução ao grande problema de programação quadrática (QP) que acontece nos demais algoritmos SVM (*Support Vector Machines*). O funcionamento de uma SVM pode ser descrito basicamente pela divisão das classes através de uma linha de separação chamada hiperplano. Tal linha busca maximizar a distância entre os pontos mais próximos em relação a cada uma das classes. Para solucionar o problema quadrático presente nas SVM o algoritmo SMO divide este problema em uma série de outros problemas QP de menor tamanho possível. O SMO possui complexidade de tempo linear para o tamanho do conjunto de treinamento.

1.3.2. Método de treinamento

Inicialmente, foi feito um teste utilizando como treinamento todos os arquivos de áudio da base em alemão *Berlin Database of Emotional Speech*[4]. Ao utilizar os arquivos de áudio em inglês da base RAVDESS[14] como entrada, os resultados não foram como esperados obtendo uma acurácia extremamente baixa, de tal forma que não houvesse a necessidade de documentar este teste. Fazendo o inverso, utilizando a base em inglês como treinamento e o alemão como entrada, os resultados continuam ruins. Por mais que a escrita das duas linguagens sejam semelhantes, a diferença da pronúncia e fonética entre elas gerou péssimos resultados. Dessa forma, foi escolhida uma outra abordagem que é o método de treinamento por Validação Cruzada e o foco em apenas um dos idiomas por vez. A Validação Cruzada[15] é uma técnica para avaliar a generalização de um conjunto de dados. Esta técnica é utilizada, principalmente, em configurações onde o objetivo é a predição e deseja-se saber quão preciso um modelo preditivo será. Neste trabalho foi utilizada a validação 10-*fold*. Em uma validação cruzada *k-fold*, a amostra de dados original é dividida em k sub-amostras iguais, uma das k sub-amostras é escolhida para teste e as $k-1$ restantes são utilizadas para treinamento, esse processo é repetido k vezes. A Figura 1.2 descreve esse processo.

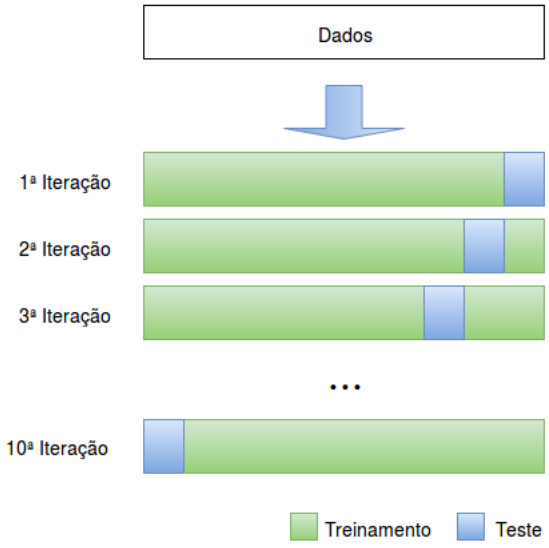

Figura 1.2. Validação Cruzada 10-fold

1.4. Experimentos e resultados

Para iniciar os experimentos, foi necessário a separação dos arquivos de cada emoção em uma pasta específica para que o *script* entenda cada classe de emoção. Após realizada a separação dos arquivos, o diretório raiz do *openSMILE* foi acessado e então direcionado para o diretório `/scripts/modeltrain`. O *script* escrito em *Perl* `makemodel.pl` foi executado, tendo como parâmetro a pasta onde se encontravam todos os arquivos de áudio e o arquivo de configuração denominado `emobase.conf`, disposto na pasta `/config`

Finalizada a execução deste comando, um diretório `/work` foi gerado, contendo um arquivo denominado `emobase` e de extensão ARFF, sendo esta a extensão utilizada para arquivos de entrada do software *Weka*. Esse arquivo e o *Weka* foram utilizados para gerar o modelo de treinamento.

Foram realizados diversos testes, porém os que tiveram melhores resultados foram utilizando a base de dados alemã EmoDB, afinal é um idioma com sons mais diversificados entre as emoções, sendo mais fácil a separação e classificação dessas. Para os testes da base de dados em inglês foi necessário restringir a apenas 4 emoções (Raiva, Felicidade, Tristeza e Surpresa) para obter um resultado relevante. As subseções abaixo descrevem os testes.

1.4.1. Experimento utilizando todas emoções da base de dados EmoDB

Nesta seção foi observado um comparativo de desempenho dos algoritmos J48 e SMO em uma base de dados alemã, a base EmoDB, que é considerada estado da arte e frequentemente utilizada em vários outros trabalhos.

Tabela 1.2. Comparativo entre as estatísticas dos algoritmos J48 e SMO utilizando 7 emoções da base EmoDB

	J48	**SMO**
Instâncias classificadas corretamente	296 (55.3271%)	468 (87.4766%)
Instâncias classificadas incorretamente	239 (44.6729%)	67 (12.5234%)
Estatística Kappa	0.472	0.8512
Erro absoluto médio	0.1297	0.2062
Raiz do erro quadrático médio	0.3426	0.3043
Erro absoluto relativo	53.7606%	85.3983%
Raiz do erro médio relativo	98.6662%	87.6486%
Número total de instâncias	535	535

Como pode ser visto na Tabela 1.2 a acurácia para o algoritmo J48 atingiu 55.35%, esse experimento obteve um resultado ruim. Já o algoritmo SMO mostrou uma acurácia de 87.47%, sendo o experimento que obteve a melhor acurácia entre todos os demais. A Tabela 1.3 descreve as estatísticas de verdadeiros positivos e falsos positivos para cada uma das emoções.

1.4.2. Experimento utilizando apenas 4 emoções da base de dados RAVDESS

Foram realizados testes com todas as emoções da base RAVDESS, mas os resultados foram muito abaixo do esperado, dessa forma optou-se por reduzir as classes

Tabela 1.3. Detalhes de acurácia para cada uma das classes de emoções utilizando as 7 emoções da base EmoDB

	J48		SMO	
Classe	**Verdadeiros Positivos**	**Falsos Positivos**	**Verdadeiros Positivos**	**Falsos Positivos**
Raiva (anger)	0.685	0.074	0.913	0.049
Tédio (boredom)	0.543	0.081	0.914	0.02
Nojo (disgust)	0.239	0.067	0.848	0.012
Medo (fear)	0.464	0.099	0.826	0.015
Felicidade (happiness)	0.493	0.099	0.746	0.024
Neutra (neutral)	0.557	0.053	0.937	0.018
Tristeza (sadness)	0.694	0.049	0.887	0.013

de emoções. O intuito do experimento desta seção foi justamente aumentar a acurácia a partir da redução da quantidade de classes de emoções. Afinal existem emoções que possuem características semelhantes entre si, por exemplo emoções de baixa excitação como: tristeza, tédio e neutra podem ser confudidas. Dessa forma foram separadas apenas 4 emoções, sendo elas: Raiva, Felicidade, Tristeza e Surpresa, isto é, duas emoções negativas (Raiva e Tristeza), uma consideravelmente neutra (Surpresa) e uma positiva (Felicidade). Tais emoções são as de maior relevância para analisar os perfis de indivíduos, baseado nos estudos futuros em questão, como por exemplo a melhoria de um sistema virtual de aprendizagem.

Tabela 1.4. Comparativo entre as estatísticas dos algoritmos J48 e SMO utilizando 4 emoções da base RAVDESS

	J48	**SMO**
Instâncias classificadas corretamente	413 (53.78%)	609 (79.2969%)
Instâncias classificadas incorretamente	355 (46.22%)	159 (20.70%)
Estatística Kappa	0.3837	0.724
Erro absoluto médio	0.2346	0.2756
Raiz do erro quadrático médio	0.4689	0.3482
Erro absoluto relativo	62.57%	73.4943%
Raiz do erro médio relativo	108.28%	80.4165%
Número total de instâncias	768	768

Observa-se pela Tabela 1.4 que a acurácia do algoritmo J48 é ainda menor do que o experimento da Tabela 1.2 mesmo tendo utilizado apenas 4 classes de emoções. Portanto, mesmo diminuindo o número de instâncias da base RAVDESS, o ganho na acurácia foi relativamente baixo. Já para o algoritmo SMO, a acurácia aumentou consideravelmente atingindo quase 80% conforme pode ser visto na Tabela 1.4. O que é uma melhoria tremenda comparado aos 53% do classificador J48. A Tabela 1.5 evidencia as estatísticas de cada classe de emoção utilizando a base RAVDESS.

De acordo com os experimentos, notou-se que o melhor classificador foi o SMO. Além disso, o melhor experimento foi utilizando todos os arquivos da base de dados EmoDB que obteve 87.47%, enquanto que na base RAVDESS foi necessário filtrá-la deixando apenas 4 emoções (Raiva, Felicidade, Tristeza e Surpresa), com isso atingindo acurácia de 79.30% que também é consideravelmente boa para o estilo de problema sendo

Tabela 1.5. Detalhes de acurácia para cada uma das classes de emoções utilizando as 4 emoções da base RAVDESS

	J48		SMO	
Classe	**Verdadeiros Positivos**	**Falsos Positivos**	**Verdadeiros Positivos**	**Falsos Positivos**
Raiva (angry)	0.557	0.16	0.818	0.075
Felicidade (happiness)	0.516	0.215	0.755	0.082
Tristeza (sadness)	0.589	0.116	0.813	0.061
Surpresa (surprised)	0.49	0.125	0.786	0.059

estudado conforme pode ser visto no trabalho de Roeland [11].

Embora a acurácia para a base em língua inglesa tenha sido um pouco menor, é bastante válido estudá-la e utilizá-la para a aplicação *web* construída, porque se trata de uma língua universal, há mais aplicações possíveis e um maior públlico poderá ser beneficiado com essa escolha. Para um projeto futuro o objetivo é a confecção de uma base em português.

1.5. Emoções online

Esta seção descreve a visão geral da aplicação web desenvolvida.

1.5.1. Visão geral do projeto

Nesse trabalho foi desenvolvida uma aplicação *Web*, que conta com o auxílio de *softwares* como *openSMILE* e *Weka* para a classificação de emoções através da voz, conforme pode ser visto na Figura 1.3

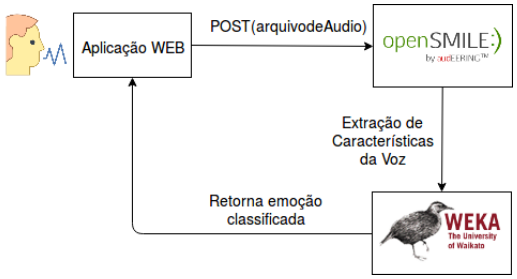

Figura 1.3. Visão geral da aplicação.

1.5.1.1. Script para análise e classificação de emoções

O *script* funciona da seguinte forma: o arquivo SMILExtract é executado recebendo o arquivo de áudio como parâmetro, após isso o classificador é executado na ferramenta Weka recebendo o modelo criado e o arquivo `output.arff` que foi gerado pelo *openSMILE*

1.5.1.2. Aplicação web

Foi utilizada a linguagem JavaScript para tratar o comportamento referente a gravação do áudio e envio dessa gravação para o *back-end* via AJAX. Para a linguagem *back-end* foi utilizado o NodeJS, que executa código Javascript no lado do servidor. A biblioteca RecordRTC foi utilizada para facilitar a gravação de áudio na página web e realizar upload para o servidor.

A aplicação respondeu razoavelmente bem aos testes executados, o grande problema é a qualidade de gravação e os possíveis ruídos que podem interferir no sinal da voz. São muitas variáveis envolvidas, como qualidade do microfone, local em que o indivíduo se encontra, a perda de qualide ao enviar o áudio para o servidor, entre outras.

1.6. Conclusão

Neste trabalho foi desenvolvida uma aplicação web capaz de gravar áudios e reconhecer emoções através destes. Para a extração de características do áudio a ferramenta *openSMILE* foi utilizada e para a classificação foi utilizado o classificador SMO providenciado pela ferarmenta Weka. Foi alcançado 79.29% de acurácia ao utilizar a base de dados RAVDESS e 87.47% ao utilizar a base EmoDB, porém essa acurácia pode variar devido a ruídos nos sinais de áudio. Essa aplicação foi concebida com o objetivo de futuramente aprimorar um ambiente virtual de aprendizagem, com o intuito de melhorar a experiência de um aluno. Para trabalhos futuros deve-se verificar possíveis melhorias em classificadores envolvendo a base RAVDESS ou outra base em inglês. Deve ser avaliado a existência de uma base em português ou a concepção desta para a realização de outros testes.

Referências

[1] R. COWIE. “Emotion recognition in human-computer interaction”. IEEE Signal processing magazine, IEEE, v.18, n.1, pp. 32–80, 2001.

[2] P. Ekman. “An argument for basic emotions”. Taylor & Francis, v.6, n. 3-4, pp. 169–200, 1992.

[3] R. Plutchik. “A general psychoevolutionary theory of emotion”. *Theories of Emotion*, Acamedic Press New York, v.1, pp. 4, 1980.

[4] F. Burkhardt. “A database of german emotional speech”. 2005, vol. 5, pp. 1517–1520.

[5] I. Murray, J. Arnott. “Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion”. *The Journal of the Acoustical Society Of America,* ASA, v.93, n.2, pp. 1097-1108, 1993.

[6] R. Iriya. “Análise de sinais de voz para reconhecimento de emoções”. Tese (Doutorado) - Universidade de São Paulo, 2014.

[7] M. Lutter. *Mel-Frequency Cepstral Coefficients*, não publicado.

[8] S. Davis, P. Mermelstein. “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”. *IEEE transactions on acoustics, speech, and signal processing,* IEEE, v.28, n.4, pp. 357-366, 1980.

[9] F. Eyben. “Recent developments in openSMILE, the munich open-source multimedia feature extractor. In: *Proc. of ACM Multimedia 2013*. Barcelona, Spain: ACM, 2013. pp. 835-838.

[10] M. Hall. “The weka data mining software: an update.” *ACM SIGKDD explorations newsletter,* ACM, v. 11, n. 1, pp. 10–18, 2009.

[11] J. Roeland. “Multisensory emotion recognition with speech and facial expression.” 2014.

[12] R. Quinlan. “C4.5: Programs form Machine Learning” San Mateo, CA: Morgan Kaufmann Publishers, 1993.

[13] J. Platt. “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vecotr Machines”, 1998.

[14] S. Livingstone, K. Peck, F. Russo. “Ravdess: The ryerson audio-visual database of emotional speech and song.” In: *Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science*, 2012.

[15] R. Kohavi. “A study of cross-validation and bootstrap for accuracy estimation and model selection.” Stanford, CA. 1995, v. 14, n. 2, pp. 1137-1145.

[16] Logan, B. (2000, October). Mel Frequency Cepstral Coefficients for Music Modeling. In ISMIR (Vol. 270, pp. 1-11).

[17] Hasan, M. R., Jamil, M., Rahman, M. G. R. M. S. (2004). Speaker identification using mel frequency cepstral coefficients. variations, 1(4).

[18] Molau, S., Pitz, M., Schluter, R., Ney, H. (2001). Computing mel-frequency cepstral coefficients on the power spectrum. In Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP’01). 2001 IEEE International Conference on (Vol. 1, pp. 73-76). IEEE.

[19] Joshy, J., Sambyo, K. (2016). A Comparison and Contrast of the Various Feature Extraction Techniques in Speaker Recognition. International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(11), 99-108.

6. Sistemas de Recomendação

Chapter 6

Implementação e avaliação de sistemas de recomendação, tradicionais e sensíveis ao contexto, baseados em técnicas de fatoração de matrizes

Vítor Rodrigues Tonon, Camila Vaccari Sundermann, Solange Oliveira Rezende
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
São Carlos, SP, Brazil

Abstract

Recommendation systems based upon matrix-factorization techniques are considered state-of-the-art algorithms, because of their good scalability and accuracy. These systems can be used in its traditional way, considering only information of users and items, as well as in its context-aware form, that also considers the context of the user. This work sought to verify if the use of matrix factorization techniques improves the accuracy of traditional and context aware recommendation systems. Experiments were performed on real datasets and the preliminary results were promising.

Resumo

Sistemas de recomendação baseados em técnicas de fatoração de matrizes são considerados o estado-da-arte na literatura da área, em razão de sua boa escalabilidade e acurácia. Esses sistemas podem ser utilizados na forma tradicional, considerando apenas informações de usuários e itens, como também na forma sensível ao contexto, considerando também o contexto do usuário. Este trabalho de conclusão de curso buscou verificar se o uso das técnicas de fatoração de matrizes melhora a acurácia dos sistemas de recomendação tradicionais e sensíveis ao contexto. Experimentos foram realizados em bases de dados reais e os resultados preliminares foram promissores.

1.1. Introdução

Vendedores eletrônicos e provedores de conteúdo desejam recomendar produtos e serviços aos seus usuários de forma a aumentar a lealdade e satisfação dos usuários. Para isso, são

utilizados os sistemas de recomendação, que, com base em informações coletadas, recomendam itens (livros, filmes, músicas, etc) de interesse dos usuários.

Os algoritmos utilizados por esses sistemas são classificados em três categorias [1]: (i) filtragem baseada em conteúdo, (ii) filtragem colaborativa, e (iii) abordagens híbridas. Na filtragem baseada em conteúdo, os itens recomendados são similares aos itens que o usuário já preferiu no passado. Na filtragem colaborativa, são identificadas relações existentes entre usuários e interdependências entre os produtos, a fim de gerar as recomendações. Nas abordagens híbridas, são combinados os métodos baseados em conteúdo e colaborativo.

Dentre essas categorias, destacam-se os algoritmos baseados em filtragem colaborativa, uma vez que superam algumas das limitações das abordagens baseadas em conteúdo [2]. Existem duas técnicas principais de filtragem colaborativa [3]: a abordagem baseada nos vizinhos mais próximos e os métodos dos fatores latentes. Os métodos dos vizinhos mais próximos fazem recomendações utilizando a vizinhança dos usuários e/ou itens do sistema. Os métodos dos fatores latentes buscam explicar as preferências dos usuários identificando características latentes nos dados; alguns dos algoritmos mais bem sucedidos que implementam esses métodos são baseados em técnicas de fatoração de matrizes.

Existem inúmeros fatores que podem influenciar a preferência de um usuário, como o contexto no qual ele está inserido [1]. Em recomendações de restaurantes, por exemplo, um usuário poderia preferir sorveterias no verão e restaurantes com comidas quentes no inverno. Pesquisadores começaram a perceber que o uso de informações adicionais, como tempo, local e outras, permite que recomendações mais precisas sejam produzidas [4, 5]. Dessa forma, um sistema de recomendação sensível ao contexto é uma tecnologia de filtragem de informação que, além do comportamento e do interesse do usuário, também utiliza essa informação contextual para recomendar itens que lhe são de interesse.

Dentre os sistemas de recomendação sensíveis ao contexto, existem métodos que também se baseiam em técnicas de fatoração de matrizes. Nesse contexto, a decomposição de tensores, que consiste em uma extensão da fatoração de matrizes tradicional, adaptada para lidar com a informação de contexto, pode ser utilizada [6]. Como os métodos tradicionais baseados em fatoração de matrizes geralmente apresentam boa escalabilidade e acurácia [3] e o uso de contexto proporciona recomendações mais precisas e úteis ao usuário [1], este trabalho buscou responder se a qualidade das recomendações melhora quando são utilizadas técnicas de fatoração de matrizes em sistemas de recomendação sensíveis ao contexto.

Para isso, os objetivos propostos foram pesquisar, explorar e implementar duas abordagens de sistemas de recomendação baseadas em fatoração de matrizes: (i) a abordagem tradicional, que não utiliza informação de contexto e (ii) uma abordagem baseada em contexto. Essas abordagens foram implementadas e inseridas no *framework* de sistemas de recomendação CARSLibrary[1]. Os experimentos conduzidos, utilizando bases de dados reais, apresentaram resultados promissores.

[1] https://github.com/maddomingues/CARSlibrary

1.2. Fundamentação Teórica

Os sistemas de recomendação com filtragem colaborativa que utilizam o método dos fatores latentes criam um modelo preditivo a partir dos dados de avaliação disponíveis, utilizando-o para fazer recomendações. A criação desse modelo permite encontrar os fatores latentes, que são padrões e características ocultas nos dados, descrevendo as características dos usuários e dos itens [3].

Alguns dos algoritmos mais bem sucedidos que implementam o método dos fatores latentes são baseados em técnicas de fatoração de matrizes. Várias pesquisas recentes [3, 7, 8, 9] têm mostrado que os algoritmos de recomendação baseados em técnicas de fatoração de matrizes apresentam bons resultados em termos de acurácia e escalabilidade. A ideia que dá suporte às técnicas de fatoração de matrizes em sistemas tradicionais é, considerando m usuários e n itens, decompor a matriz de utilidade R, que armazena as preferências dos usuários em relação aos itens, em duas outras matrizes, X e Y. Cada uma dessas matrizes possui k fatores, que são características inferidas dos dados de avaliação existentes (Equação 1).

$$\hat{R} = XY^T \tag{1}$$

Assim, para estimar a preferência de um usuário u para um item i, tem-se que:

$$\hat{r}_{u,i} = \vec{x}_u \cdot (\vec{y}_i)^T \tag{2}$$

tal que $\vec{x}_u$ corresponde ao vetor de fatores do usuário u e $\vec{y}_i$, ao vetor de fatores do item i. Essas informações podem ser obtidas, respectivamente, das matrizes X e Y.

Para obter os vetores de fatores ($\vec{x}_u$ e $\vec{y}_i$), o sistema de recomendação deve minimizar a raiz quadrada do erro quadrático médio (RMSE) no conjunto de avaliações conhecidas [3]:

$$\min_{x,y} \sum_{(u,i)\in K} (r_{u,i} - \vec{x}_u \cdot (\vec{y}_i)^T)^2 + \lambda(\|\vec{x}_u\|^2 + \|\vec{y}_i\|^2) \tag{3}$$

tal que K é o conjunto dos pares (u,i) para os quais $r_{u,i}$ é conhecido. A constante λ é utilizada para a regularização dos parâmetros aprendidos, de forma a evitar a superespecialização do modelo.

Nos sistemas de recomendação sensíveis ao contexto também é possível utilizar técnicas de fatoração de matrizes para gerar as recomendações. Como esses sistemas trabalham com espaços multidimensionais, uma nova estratégia deve ser considerada. Para isso, são utilizados os tensores que, por serem *arrays* multidimensionais, permitem representar todas as dimensões do problema de recomendação ($Usuário \times Item \times Contexto$). Dessa forma, a matriz de utilidade que faz parte da recomendação tradicional transforma-se em um tensor de utilidade. A decomposição desses tensores consiste de uma extensão da fatoração de matrizes tradicional, adaptada para lidar com informações n dimensionais [6].

Existem múltiplas formas de decompor tensores, tais como: Decomposição de Tucker [10] e Decomposição Canônica [11]. O primeiro método permite decompor um tensor de ordem N em N matrizes e um outro tensor, de tamanho inferior ao do original. O outro método, por sua vez, decompõe um tensor como uma soma de outros tensores

mais simples. No entanto, algumas adaptações devem ser feitas para utilizar essas abordagens no domínio dos sistemas de recomendação sensíveis ao contexto, uma vez que esses métodos não são apropriados para lidar com dados de entrada esparsos, como no caso dos sistemas de recomendação. Nesse contexto, em [6], foi proposto um modelo genérico que utiliza a Decomposição de Tucker para a geração das recomendações, utilizando apenas os dados observados do tensor de utilidade. Essa abordagem, apesar de apresentar bons resultados, possui alta complexidade, o que dificulta a sua utilização em cenários de recomendação reais. Em [12], foi proposto o algoritmo CP-WOPT, que modela a recomendação como um problema de mínimos quadrados ponderados, utilizando apenas os dados conhecidos e a Decomposição Canônica. Nos experimentos realizados em [12], foi observado que o algoritmo CP-WOPT é capaz de fatorar os tensores de maneira eficaz mesmo quando os dados disponíveis são escassos. Esses modelos, apesar de apresentarem bons resultados, são relativamente complexos; existem outras maneiras mais simples de se aplicar os princípios dos fatores latentes no cenário multidimensional.

Nesse contexto, em [13], é descrito um método que realiza interações par-a-par entre as diferentes dimensões, simplificando o processo. Nessa abordagem, o tensor de utilidade R, com dimensionalidade n, é decomposto em n matrizes de fatores: $U_1, U_2, U_3, \cdots, U_n$, tal que as matrizes U_1 e U_2 representam os fatores de usuários e itens respectivamente, e a matriz U_a $(3 \leq a \leq n)$ representa a matriz de fatores da variável de contexto a. O princípio básico desta abordagem é realizar interações par-a-par entre as diferentes matrizes, possibilitando predizer o *rating* da forma como se segue:

$$\hat{r}_{i_1,\cdots,i_n} = \sum_{a<b\leq n} [U_a (U_b)^T]_{i_a i_b} \tag{4}$$

em que i_j é um possível valor para a variável j.

As matrizes de fatores U_a são de tamanho $s_a \times k$, em que s_a representa a quantidade de valores distintos que a variável a assume e k representa o número de fatores. Para a sua obtenção, o sistema deve minimizar a seguinte função:

$$J = \sum_{(i_1,\cdots,i_n)\in S} (r_{i_1,\cdots,i_n} - \sum_{a<b\leq n} [U_a (U_b)^T]_{i_a i_b})^2 + \lambda \sum_{a=1}^{n} ||U_a||^2 \tag{5}$$

tal que S é o conjunto de entradas $(i_1, \cdots, i_n)$ para os quais $r_{i_1,\cdots,i_n}$ é especificado. Da mesma forma como na recomendação tradicional, o parâmetro λ é usado para a regularização do modelo, evitando a sua superespecialização.

1.3. Materiais e Métodos

Os métodos de recomendação baseados em técnicas de fatoração de matrizes, descritos na Seção 1.2, foram implementados na linguagem de programação R. Posteriormente, foram inseridos no *framework* de sistemas de recomendação CARSLibrary, que disponibiliza implementações para os seguintes algoritmos de recomendação tradicional e sensíveis ao contexto: *IBCF* [14], *cReduction* [1], *daviBEST* [15], *weightPoF* e *filterPoF* [4]. Este *framework* considera avaliações unárias/binárias, isto é, valores um ou zero que indicam se o usuário acessou/avaliou o item ou não.

Conforme os algoritmos são executados na CARSLibrary, também é realizada a avaliação *offline*, com *10-fold cross validation*, utilizando o protocolo *All But One* [16]. As métricas implementadas no *framework* são: *Recall*, *Precision*, *F1*, *Fallout* e *MAP*.

Para permitir a integração dos algoritmos implementados neste projeto ao *framework*, foi necessário desenvolver um conjunto de módulos para cada algoritmo: **Módulo de Leitura**, **Módulo de Recomendação** e **Módulo de Avaliação**. Na Figura 1.1, é apresentada a estrutura proposta para cada algoritmo implementado, ilustrando como esses módulos interagem com o sistema.

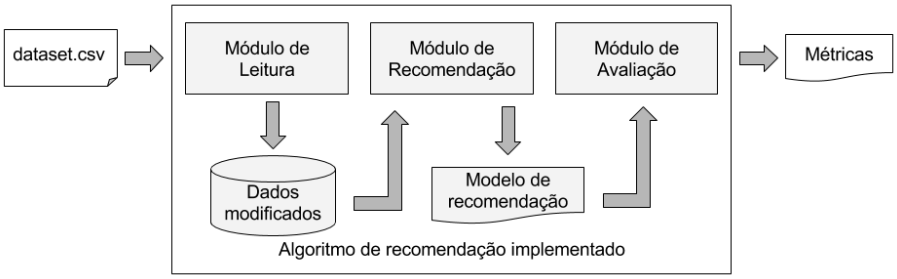

Figure 1.1: Módulos desenvolvidos para a integração dos algoritmos implementados ao *framework* CARSLibrary

Enquanto o *framework* trabalha com dados de avaliação unários/binários, os algoritmos de recomendação utilizados neste trabalho consideram dados de avaliação numéricos, que indicam o grau de preferência que o usuário possui por esse item. Como o arquivo de entrada para o *framework* contém apenas esses dados unários, é necessário adaptá-lo para o algoritmo de recomendação implementado. Nesse contexto, o **Módulo de Leitura** lê o arquivo e infere o grau de preferência para cada par usuário-item nele contido, gerando um novo conjunto de dados modificados. Para isso, é considerado o número de vezes que um usuário acessou um dado item. Afinal, quanto maior o número de acessos que um dado usuário fez a uma página *Web*, por exemplo, maior a probabilidade desta página conter algum conteúdo de seu interesse. Assim, se no arquivo de entrada contiverem três acessos do usuário u ao item i, por exemplo, considera-se que o usuário u viu o item i com um grau de preferência 3.

Os dados modificados são usados como entrada do **Módulo de Recomendação**, que implementa os dois algoritmos de recomendação considerados neste projeto, que foram descritos na Seção 1.2. A implementação consiste em minimizar as Equações 3 e 5, referentes ao algoritmo tradicional (*Matrix Factorization* - MF) e ao sensível ao contexto (*Context-Aware Matrix Factorization* - CAMF), por meio da técnica *Stochastic Gradient Descent* (SGD) [17]. Essa técnica, adaptada para o domínio de sistemas de recomendação, itera sobre todos os dados de avaliação, realizando o treinamento do modelo de recomendação, que é gerado como saída desta etapa.

O modelo de recomendação produzido é então utilizado pelo **Módulo de Avaliação** para gerar as recomendações para cada usuário presente na base. Utilizando essas recomendações e os dados de teste, a avaliação *offline* é realizada e as métricas consideradas pelo *framework* são retornadas.

1.4. Experimentos e Resultados

Os experimentos realizados tiveram como objetivo verificar se o uso da fatoração de matrizes aumenta a acurácia dos sistemas de recomendação, tanto tradicionais quanto sensíveis ao contexto.

Foram utilizadas duas bases de dados para a avaliação dos sistemas: (i) base de dados que se refere aos acessos feitos por usuários no site da Agência Embrapa de Informação Tecnológica, pertencente à Embrapa Informática Agropecuária - CNPTIA, no período de novembro de 2012; e (ii) base de dados, disponibilizado pela empresa Onion Tecnologia Inteligente[2], que se refere aos produtos comprados por usuários em restaurantes ao utilizarem o aplicativo de cardápio inteligente *Onion Menu*.

A primeira base de dados foi utilizada em um trabalho de mestrado [5]. Nessa ocasião, os *logs* de acessos às páginas da Embrapa foram compilados para utilização em um método de extração automática de informações contextuais. Em seus experimentos, foram geradas duas bases de dados derivadas, acrescidas da informação contextual, que correspondem aos tópicos obtidos a partir de uma hierarquia de tópicos construída utilizando o método LIHC [18]. Essas novas bases de dados, que diferem apenas na granularidade utilizada para a seleção dos tópicos considerados, são referidas como `Embrapa A` e `Embrapa B` no restante deste artigo.

Na segunda base de dados foram consideradas as informações de 4304 pedidos realizados, considerando 470 usuários e 22 restaurantes do sistema, acrecidas das seguintes informações de contexto referentes aos pedidos feitos no aplicativo: *dia da semana*, *período do dia*, *tipo do serviço* e *distância*. Nos experimentos, foram utilizadas duas bases de dados derivadas: uma acrescida das quatro informações contextuais e outra acrescida de apenas da informação de *distância*. Essas novas bases de dados serão referenciadas como `Onion C` e `Onion D`, respectivamente.

Primeiramente, foi necessário determinar a combinação dos parâmetros λ e k que gera o melhor resultado para cada base de dados. Resultados preliminares indicaram que as bases `Embrapa A` e `Embrapa B` apresentaram melhores resultados quando foram utilizados os parâmetros $\lambda = 200$ e $k = 10$. As bases `Onion C` e `Onion D`, por sua vez, apresentaram melhores resultados para os parâmetros $\lambda = 1$ e $k = 5$.

Neste trabalho, para analisar o desempenho dos algoritmos de recomendação implementados, foram realizados experimentos de forma *offline*, utilizando o Protocolo *All But One* [16], com *10-fold cross validation* e a métrica MAP@N como medida de desempenho (com N igual a 10 recomendações). Os valores de MAP nos 10 *folds* foram sumarizados usando a média. Além disso, foram utilizados como *baseline* os seguintes algoritmos já implementados no *framework* CARSLibrary: *Item-Based Collaborative Filtering* para todos os experimentos; *cReduction*, *weightPoF* e *filterPoF* para as bases de dados `Embrapa A` e `Embrapa B`, visto que apresentaram os melhores resultados nessas bases em experimentos anteriores; e *weightPoF* e *daviBEST* para as bases de dados da *Onion Menu*, que também apresentaram bons resultados em outros experimentos. Por fim, para comparar dois algoritmos de recomendação, foi aplicado o teste t-student pareado com 95% de nível de confiança.

[2]http://www.onionapp.com.br/

Na Figura 1.2, é exibido um gráfico comparando a métrica de desempenho MAP@10 obtida pelos algoritmos de recomendação implementados neste projeto (MF e CAMF) com a mesma métrica obtida pelos *baselines* considerados.

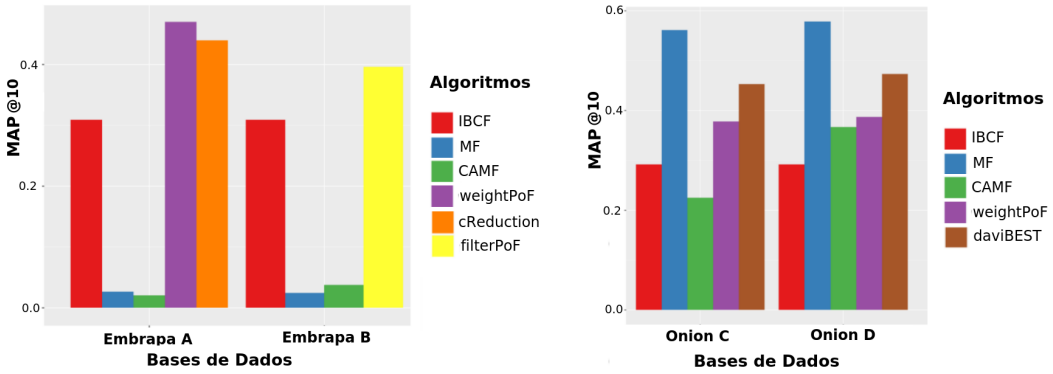

Figure 1.2: Comparação do desempenho dos algoritmos em relação à métrica MAP@10

Na base de dados `Embrapa A`, tanto o algoritmo de fatoração de matrizes tradicional (MF) quanto o algoritmo sensível ao contexto (CAMF) implementados apresentaram resultados estatisticamente inferiores aos *baselines* considerados. Além disso, o algoritmo sensível ao contexto gerou recomendações inferiores quando comparadas às recomendações do algoritmo tradicional baseado em fatoração de matrizes.

Da mesma forma, na base de dados `Embrapa B`, os dois algoritmos de fatoração de matrizes utilizados, MF e CAMF, apresentaram resultados estatisticamente inferiores do que os seus *baselines*, IBCF e *filterPoF*, respectivamente. No entanto, quando os algoritmos MF e CAMF são comparados entre si, observa-se que apresentam resultados similares.

Na base de dados `Onion C`, o algoritmo de recomendação tradicional baseado em fatoração de matrizes implementado (MF) apresentou resultados estatisticamente superiores ao *baseline* IBCF. Ainda assim, o algoritmo sensível ao contexto (CAMF) não conseguiu superar os algoritmos *daviBEST* e *weightPoF*, que também utilizam contexto.

Por fim, na base de dados `Onion D`, o algoritmo de recomendação tradicional MF apresentou resultados estatisticamente superiores ao IBCF. Apesar de não ter superado o algoritmo tradicional (MF), o algoritmo sensível ao contexto implementado (CAMF) apresentou resultados estatisticamente iguais aos *baselines* sensíveis ao contexto utilizados (*daviBEST* e *weightPoF*).

De maneira geral, os algoritmos implementados apresentaram resultados não tão satisfatórios para as bases de dados `Embrapa A` e `Embrapa B`. No entanto, para a base de dados `Onion C` e `Onion D`, os resultados obtidos pelo algoritmo de recomendação tradicional baseado em fatoração de matrizes foram muito bons, superando tanto o *baseline* tradicional quanto os *baselines* que utilizam contexto. Porém, na base de dados `Onion C`, o algoritmo sensível ao contexto implementado (CAMF) não conseguiu superar os *baselines* sensíveis ao contexto considerados e na base de dados `Onion D` os resultados foram similares aos dos *baselines*.

Alguns dos resultados insatisfatórios podem ser explicados pela abordagem escolhida para a implementação do Módulo de Leitura, descrito na Seção 1.3. Como os dados de avaliação disponíveis são utilizados para a inferência da preferência do usuário (a quantidade de avaliações para um dado par usuário-item é usada como entrada do sistema de recomendação), o desempenho do sistema de recomendação pode ser afetado quando não há avaliações suficientes para os usuários-itens dessas bases ou quando os itens possuem aproximadamente o mesmo número de avaliações. Isso pode dificultar a identificação das preferências do usuários pelos sistemas de recomendação. Para solucionar esse problema, é necessário garantir que existam dados de avaliação suficientes para cada usuário/item nos dados de treinamento.

1.5. Conclusões

O objetivo do trabalho foi verificar se os algoritmos baseados em técnicas de fatoração de matrizes apresentam resultados superiores às técnicas tradicionalmente encontradas na literatura, que são, em geral, métodos de recomendação baseados em vizinhança. Além disso, buscou-se verificar se o uso do contexto em sistemas que utilizam fatoração de matrizes possibilita a geração de recomendações mais precisas.

Os experimentos realizados mostraram que o algoritmo MF conseguiu superar o *baseline* tradicional para duas das bases de dados consideradas. Além disso, os resultados do algoritmo CAMF mostram que o algoritmo não conseguiu superar o desempenho dos *baselines* sensíveis ao contexto considerados. No entanto, esses resultados insatisfatórios podem estar relacionados com a abordagem escolhida para a implementação do Módulo de Leitura. Diante disso, como trabalhos futuros, pretende-se realizar novos experimentos utilizando bases de dados com maior número de avaliações, para avaliar se o algoritmo CAMF apresenta resultados superiores aos obtidos nos experimentos realizados nesse trabalho. Além disso, outras abordagens de recomendação sensíveis ao contexto baseadas em fatoração de matrizes podem ser exploradas, a fim de comparar o seu desempenho ao algoritmo sensível ao contexto implementado neste trabalho (CAMF).

As principais contribuições deste trabalho são os algoritmos de recomendação baseados em fatoração de matrizes implementados, que apresentaram resultados promissores, mesmo que ainda sejam necessários novos experimentos, e a sua integração ao *framework* CARSLibrary, que possibilita a sua utilização pela comunidade.

1.6. Agradecimentos

Este projeto foi desenvolvido com o apoio da CAPES e da Fundação de Amparo à Pesquisa do Estado de São Paulo (Processos FAPESP nº 2016/25075-1 e nº 2018/04651-0)

References

[1] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating contextual information in recommender systems using a multidimensional approach," *ACM Transactions on Information Systems (TOIS)*, vol. 23, no. 1, pp. 103–145, 2005.

[2] C. Desrosiers and G. Karypis, "A comprehensive survey of neighborhood-based rec-

ommendation methods," in *Recommender systems handbook*. Springer, 2011, pp. 107–144.

[3] Y. Koren, R. Bell, C. Volinsky *et al.*, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[4] U. Panniello and M. Gorgoglione, "Incorporating context into recommender systems: an empirical comparison of context-based approaches," *Electronic Commerce Research*, vol. 12, no. 1, pp. 1–30, 2012.

[5] C. V. Sundermann, M. A. Domingues, M. da Silva Conrado, and S. O. Rezende, "Privileged contextual information for context-aware recommender systems," *Expert Systems with Applications*, vol. 57, pp. 139–158, 2016.

[6] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering," in *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 2010, pp. 79–86.

[7] B. Abdollahi and O. Nasraoui, "Using explainability for constrained matrix factorization," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 2017, pp. 79–83.

[8] C. Wang, Q. Liu, R.-z. Wu, E. Chen, C. Liu, X. Huang, and Z. Huang, "Confidence-aware matrix factorization for recommender systems." in *AAAI*, 2018.

[9] H. Wu, Z. Zhang, K. Yue, B. Zhang, J. He, and L. Sun, "Dual-regularized matrix factorization with deep neural networks for recommender systems," *Knowledge-Based Systems*, vol. 145, pp. 46–58, 2018.

[10] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[11] A. Stegeman and N. D. Sidiropoulos, "On kruskal's uniqueness condition for the candecomp/parafac decomposition," *Linear Algebra and its applications*, vol. 420, no. 2-3, pp. 540–552, 2007.

[12] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, 2011.

[13] C. C. Aggarwal, *Recommender Systems*, 1st ed. Springer, 2016.

[14] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*. ACM, 2001, pp. 285–295.

[15] M. A. Domingues, A. M. Jorge, and C. Soares, "Dimensions as virtual items: Improving the predictive ability of top-n recommender systems," *Information Processing & Management*, vol. 49, no. 3, pp. 698–720, 2013.

[16] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1998, pp. 43–52.

[17] L. Bottou, "Stochastic gradient descent tricks," *Neural networks: Tricks of the trade*, pp. 421–436, 2012.

[18] R. M. Marcacini and S. O. Rezende, "Incremental hierarchical text clustering with privileged information," in *Proceedings of the 2013 ACM symposium on Document engineering*. ACM, 2013, pp. 231–232.

7. Jogo para Ensino da Arquitetura TCP/IP

Chapter 7

Jogo como Ferramenta para apoio ao Ensino/Aprendizagem da Arquitetura TCP/IP

Áquilla O. Faria Nascimento, Joselice Ferreira Lima, Adriano Antunes Prates

Abstract

Games are used in education as a means to help in the learning/teaching proccess, because of that, this article, has the goal to demonstrate an educational game called EducaRedes aimed in supporting the teaching of the TCP/IP archtecture, with TI students as his target audience. The adopted method was the litterature revision, is used the SCRUM4Games methodology to develop the game and the self report scale to analise his viability As a result, there is an digital educational game, as a tool to auxiliate in the teaching/learning proccess and the tests indicated that the contentes of the game correspond to those seeing in class, the layout is clean and uncomplicated and the game has a médium difficulty.

Resumo

Os jogos são utilizados na educação como meio para auxiliar no processo de ensino/aprendizagem. O objetivo desse artigo é demonstrar um jogo educacional denominado EducaRedes focado no ensino da arquitetura TCP/IP, tendo como público alvo acadêmicos da área da Tecnologia da Informação do ensino superior. O método adotado é a revisão de literatura com a posterior análise de cinco engines e a seleção da Cocos2D-X para desenvolver esse trabalho. São utilizadas a metodologia SCRUM4Games para o desenvolvimento do jogo e as escalas de autorrelato para a análise de sua viabilidade. Como resultados, concluiu-se a criação do jogo e a análise de viabilidade indica que o conteúdo deste corresponde à disciplina, o layout é compreensível e descomplicado e o jogo possui dificuldade mediana.

1.1. Introdução

A Redes de computadores são um conjunto de computadores autônomos interconectados por uma tecnologia de comunicação. É a interligação de dispositivos através de uma rede de transmissão de dados[1].A popularidade das redes de

computadores tem aumentado, seja pelo advento da Internet ou pelo avanço das tecnologias de comunicação e transmissão de dados que possibilita o uso de smartphones abrindo oportunidade de uso em diversos espaços

Das tecnologias existentes em redes de computadores, uma das mais proeminentes é a arquitetura TCP/IP (*Transfer Control Protocol/Internet Protocol)*. TCP/IP é um conjunto de protocolos que permite que computadores, smartphones e sistemas embarcados de todos os tamanhos, fornecedores e sistemas comuniquem entre si. A arquitetura TCP/IP compreende um conjunto de serviços oferecidos por camadas para que haja a comunicação entre diversos dispositivos na rede[2].

Entende-se que o TCP/IP é a base para a rede mundial de computadores e os protocolos desta arquitetura possuem regras que devem ser seguidas para permitir a comunicação entre dispositivos localizados remotamente. Entretanto o seu entendimento e aprendizado nos cursos de engenharia e computação é enfadonho, cuja alternativa é utilizar de recursos lúdicos, tais como jogos.

Nesse contexto, "todo jogo verdadeiro é uma metáfora da vida. Logo, representa, de uma forma fiel ou não, elementos presentes no cotidiano popular"[3]. Quando utilizados nas atividades de ensino, podem propiciar momentos lúdicos e interativos no processo de aprendizagem[4].

Diante deste cenário, este artigo busca alternativas para o ensino de redes de computadores, uma vez que existe uma carência de material didático lúdicos focados no ensino da arquitetura TCP/IP na bibliografia analisada.

O objetivo do artigo é apresentar um jogo denominado EducaRedes visando motivar e fixar os conhecimentos sobre os conteúdos de redes de computadores. Avaliou-se o jogo proposto em duas turmas de um curso superior na área de informática e o resultado apontou que o jogo é desafiador, bem explicativo e com o conteúdo a par com o que é visto em sala de aula.

1.2. Fundamentação Teórica

1.2.1. Rede de computadores

Uma rede de computadores é uma coleção de computadores e outros dispositivos (nós) que usam um protocolo comum para compartilhar recursos uns com os outros através do auxílio do meio da rede [1].

As redes de computadores surgiram da necessidade da troca de informações, já que é possível ter acesso a um dado que está fisicamente localizado distante do indivíduo[5].

Para que se tenha acesso a essa informação é necessário que uma série de serviços, protocolos e camadas da rede funcionem. Os protocolos, são conjunto das informações, normas e regras definidas a partir de um ato oficial, são o que regem as comunicações intra-computadores. Um protocolo de comunicação de rede é um conjunto estabelecido ou aceito de procedimentos, regras ou especificações formais que governam um comportamento específicos na rede[6].

Muitos desses protocolos, tais quais, o *Ethernet*, TCP e o IP e serviços surgiram usando como padrão as camadas estabelecidas no modelo lógico OSI (*Open System Interconection* – Sistema de Interconexão aberto).

Do modelo lógico OSI veio a arquitetura TCP/IP que é o padrão de arquitetura para comunicações via *Internet*. Porém, tanto o modelo OSI quanto a arquitetura TCP/IP possuem similaridades e diferenças. Os dois se baseiam no conceito de uma pilha de protocolos independentes. Além disso, as camadas têm praticamente as mesmas funções[1].

As camadas da arquitetura TCP/IP são segmentadas, cada camada tem por base a sua antecessora. Os serviços e funções de cada uma variam com cada tipo de rede. No entanto, em qualquer rede, a função de cada camada é fornecer serviços para as superiores, deixando claro como esses serviços são realizados.

A arquitetura TCP/IP possui 04 (quatro) camadas, sendo elas: Acesso à rede, Internet, Transporte e Aplicação, descritas na tabela 1.

Tabela 1.1. Camadas da arquitetura TCP/IP

Camada	Descrição
Camada de Acesso à Rede	A camada tem por objetivo gerar e transmitir o sinal pelo meio, seja através da tecnologia Ethernet, Wi-fi, Bluetooth, entre outros. Um dos serviços dessa camada é a modulação da informação através de um sinal analógico ou digital.
Camada de Internet	A camada Internet tem como objetivo gerenciar pacotes na rede, identificando sua origem/destino e redes para entrega. Um serviço oferecido é o roteamento de pacotes, através dos endereços de origem/destino.
Camada de Transporte	A camada de transporte estabelece uma conexão fim a fim (conexão confiável) entre a origem e o destino dos dados. Serviço oferecido é a identificação dos saltos dados na rede
Camada de Aplicação	A camada de aplicação é responsável pelos protocolos de comunicação com as diferentes aplicações e é a interface entre os programas e a arquitetura TCP/IP. Serviço oferecido é o recebimento de arquivos HTTP (HyperText Transfer Protocol) para a exibição de páginas Web.

As disciplinas de redes de computadores, ministradas nos cursos superiores de informática possuem uma dificuldade acentuada por uma série de fatores, tais quais: dificuldade de manutenção dos laboratórios, conteúdo extenso e escassez de material interativo ou prático para que o aluno exercite aquilo que é demonstrado em sala de aula, dificuldade de aprendizado do conteúdo e parte teórica bastante extensa.

Contudo, o estudo da arquitetura TCP/IP se vê engessado devido aos desafios incutidos no ensino da disciplina nos cursos superiores, seja pela falta de laboratórios práticos, seja pela falta de conteúdo didático interativo sobre o tema.

Os investimentos teriam que ser mais amplos nas instituições de ensino de Tecnologia da Informação, ou seja, os investimentos teriam que contemplar diversas áreas da TI, tendo em vista a responsabilidade de formar profissionais qualificados e atualizados de acordo com as novas competências exigidas pelo mercado[7]. Sem a possibilidade de experimentação, muitos dos conceitos aprendidos em sala de aula acabam se perdendo na memória do aluno.

Percebe-se que há uma dificuldade de motivar o aluno no aprendizado de redes de computadores baseado só em teorias, pois os livros e artigos podem fornecer uma

boa base teórica, porém o aluno precisa comprovar na prática o conteúdo teórico, de forma que a aula fique mais interativa e o ensino menos complexo e cansativo[8].

Um dos maiores desafios enfrentados pelo professor, é a dificuldade de preparar o aluno para a prática profissional em um ambiente acadêmico, para tanto, cabe ao professor encontrar um meio de realizar atividades práticas que despertem o interesse do educando, ao invés de ministrar apenas aulas teóricas[9].

O ensino de redes de computadores já é um desafio, devido à quantidade de disciplinas que envolvem um número significativo de conceitos puramente técnicos. Dessa forma, o professor pode introduzir os conceitos por meio do ensino teórico, mas depende da aplicação prática desses conceitos para consolidar o aprendizado[7].

Com um conteúdo complexo e muitas vezes não atrativo, o aluno acaba por não absorver totalmente os conceitos e práticas passados na disciplinas de redes de computadores, muitos desses conceitos diretamente ligados à um completo entendimento da disciplina.

Nesse contexto, percebe-se que as disciplinas de redes de computadores podem utilizar-se de jogos educacionais digitais como um método de ensino/aprendizado que facilite a absorção desses conceitos.

1.2.2. Jogos Educacionais

Jogos educacionais são qualquer atividade de formato instrucional ou de aprendizagem e que seja regulado por regras e restrições[10].

Os jogos educacionais podem e são utilizados no ensino das mais diversas disciplinas. Baseiam-se numa abordagem autodirigida, isto é, aquela em que o sujeito aprende por si só, através da descoberta de relações e da interação com o software[11].

Os jogos educacionais são capazes de unir diversão e aprendizado de uma maneira estimulante, principalmente para aqueles que possuem em sua rotina, contato com diversos recursos tecnológicos atuais. Ainda segundo esse autor, que defende a não passividade do educando em seu processo de aprendizagem, os métodos de ensino tradicionais (quadro negro e giz) não possuem sua potencialidade explorada, nem mesmo através da educação online, que permite uma interação maior entre educador e educando[12].

O uso de jogos eletrônicos como ferramenta de suporte ao ensino dos conteúdo de redes de computadores é útil pois permite aumentar o potencial de interação entre aprendiz e objeto. Além disso, motiva os participantes através de desafios lúdicos[13].

Jogos educacionais abordam os mais diversos assuntos, porém, quando se trata de rede de computadores o número de jogos ou estudos sobre os mesmos nessa área é bastante reduzido.

Na revisão bibliográfica feita utilizando como bases de dados científicos os agregadores de trabalhos acadêmicos: Google Acadêmico, Base de periódicos da CAPES, sciELO, PubMed, SBGames no período compreendido entre 2010 até o ano de 2016, com o intuito de identificar na área acadêmica publicações de artigos referentes a jogos. Foram coletados artigos nas mais diversas áreas do evento, que incluem arte e design, cultura, indústria, entre outras na área de jogos. Porém, poucos artigos referentes a jogos focados ao ensino de Rede de Computadores.

Identifica-se então que o ensino da disciplina de redes de computadores via jogo educacional, pode gerar no educando uma série de fatores que façam com que o mesmo sinta vontade de aprender e entender aquele conteúdo.

1.3. Trabalhos Correlatos

Dos trabalhos encontrados que tratam de rede de computadores, pode-se citar o uso de simuladores no auxílio do ensino de redes de computadores apontando as dificuldades encontradas como falta de materiais práticos e laboratórios[7]. Identificou também as contribuições do uso de simuladores de redes como estratégias no processo de ensino-aprendizagem do conteúdo, verificando que a ferramenta que mais se adequava ao uso de simuladores de redes no ensino era a Cisco Packet Tracer, que é um programa educacional gratuito que permite simular uma rede de computadores, através de equipamentos e configurações presente em situações reais, disponível para *Windows* e *Linux* e com atualizações frequentes.

Em [14] apresenta a análise e modelagem de tráfego de jogos FPS(*First Person Shooter* – tiro em primeira pessoa) em redes 802.11 em ambientes indoor, no qual cria-se um mecanismo de análise do tráfego gerado por esse tipo de aplicação. A ideia central do autor [14] é garantir um QoS (*Quality of Service* – Qualidade de serviço) que é a capacidade de melhorar os serviços trafegados na rede sobre tecnologias de comunicação para jogos em rede 802.11. Percebeu-se nesse trabalho que poucos jogadores não degradam o desempenho da rede e aplicações tais como FTP são mais sensíveis à concorrência do jogo que outras, como HTTP.

Observa-se nos trabalhos correlatos que o ensino é facilitado com o uso de simuladores e o tráfego de redes não sofre uma interferência muito alta com jogos digitais. Contudo, na bibliografia analisada percebeu-se uma carência referente a jogos educacionais, o *CyberCIEGE* que é focados no ensino de redes de computadores, sendo um jogo sério, que são definidos como uma atividade lúdica realizada no contexto de uma realidade simulada, no qual os participantes tentam alcançar, pelo menos uma meta arbitrária, não trivial para auxiliar no ensino-aprendizado da disciplina[15]. O *CyberCIEGE* é um videogame considerado inovador para ensinar conceitos de segurança de computadores e de rede[16].

1.4. Metodologia

O artigo classifica-se quanto à natureza da pesquisa aplicada; quantitativa, bibliográfica e tem por objetivo gerar novos conhecimentos resultantes do processo de pesquisa – "como fazer"[17].

O jogo foi desenvolvido levando em conta o seu GDD (*Game Design Document*) que é o documento padrão com as especificações de um jogo digital, tais quais: História, Personagens, Gameplay, Controles e Universo do jogo. A plataforma de desenvolvimento utilizada foi A *Cocos2D-X* é um *Framework Open Source*, gratuito, com a missão de simplificar o desenvolvimento de jogos 2D. Utilizando-se de linguagens como *Javascript, C++ e LUA*, a ferramenta consegue exportar o código gerado para as mais diversas plataformas, tais quais, *Windows, Linux, IOS, Android*, entre outros.

Com relação às artes e elementos presentes no jogo, o *character design(Design* de personagem) gera os personagens que virão a aparecer no jogo, seja através de um

diálogo ou uma animação. Possui uma natureza iterativa já que para se chegar ao *design* final primeiro é necessário refinar o personagem em sua forma bruta, também conhecida como *sketch(*rascunho) todos desenvolvidos pelo autor. Já as músicas utilizadas no jogo, desenvolvidas pelo autor, foram criadas utilizando-se o programa *Fruity Loops Studio*, fazendo parte do *design* de áudio e tem como objetivo identificar diante do jogo, quais elementos, emoções e estratégias o jogador deve sentir e traduzir isso de forma sonora.

O Jogo possui quatro fases e em cada fase possui três módulos: um módulo contendo a explicação do professor referente àquela fase, um módulo contendo um minijogo explicando um conceito da arquitetura TCP/IP e por fim, um quiz com perguntas referentes ao que foi visto naquela fase.

Os testes do jogo Educaredes Versão 1.0 foi avaliado por um grupo de alunos do curso superior em Análise e Desenvolvimento de Sistemas (TADS) da disciplina de Redes de Computadores. Os resultados apresentados na seção de Resultados.

1.5. Jogo Digital Educacional EducaRedes

Em cada fase do jogo há um minijogo diferente explicando o seu conceito.

Na primeira fase é introduzido conceitos da camada de Acesso a Rede, onde é demonstrado modulação de sinais (analógico e digital) e o aluno/jogador jogará um minijogo exemplificando a geração deste sinal.

Na segunda fase é apresentado os conceitos da camada Internet e então mais uma vez o jogador é levado a um minijogo, onde os conceitos de roteamento de pacotes(origem/destino) são demonstrados.

Na terceira fase o jogador aprende-se os conceitos da camada de transporte, seus serviços e protocolos, por fim, um minijogo vem para solidificar os conceitos de saltos na rede, roteamento e como ele é realizado. Na quarta e última fase, o jogador aprende os conceitos da última camada da arquitetura TCP/IP a camada de aplicação.

A figura 1 mostra a tela inicial e seleção de fases do jogo.

Figure 1.1. **Tela Inicial e Tela de seleção de fases. EducaRedes (2018)**

No final faz uma pequena revisão do que foi visto até então e exemplifica os serviços dessa camada, no fim, o jogador testa o conteúdo absorvido em um minijogo sobre Sockets IP.

A figura 2 mostra o jogador dentro da fase de Acesso a rede e todas suas etapas:

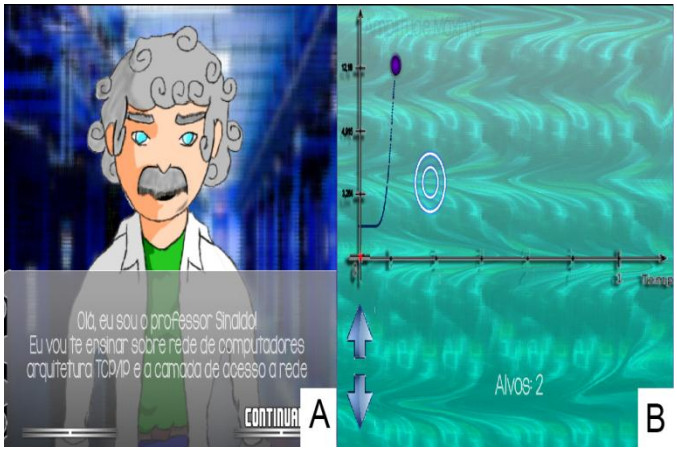

Figure 1.2. **Módulos da Fase de Acesso a rede do jogo EducaRedes (2018)**

A etapa final do desenvolvimento consistiu em avaliar a utilização do jogo, usando a escala de autorrelato. proposta por Aguiar & Soares, que se baseia nas escalas de diferencial semântico de Likert, usando perguntas com itens-likert como opções.

1.6. Testes e Avaliação

Para identificar e avaliar a utilização do jogo, usou-se a escala de autorrelato proposta por Aguiar & Soares, [18], que se baseia nas escalas de diferencial semântico de Likert, usando perguntas com itens-likert como opções.

O universo da pesquisa foram os acadêmicos do curso de Tecnologia em Análise e Desenvolvimento de Sistemas e a amostra e população foi composta por 30 (trinta) acadêmicos que cursam ou estão cursando a disciplina de Redes I e II, no Instituto Federal do Norte de Minas Gerais – Campus Januária, onde se aplicou a pesquisa.

O questionário foi dividido de acordo com as fases presentes no jogo, tendo perguntas que contemplavam seu conteúdo, layout, dificuldade, desafio, diversão e o estímulo à aprendizagem.

A coleta de dados foi feita através de um questionário estruturado. A pesquisa de campo foi executada no período de dezembro de 2017, com questões de múltipla escolha com seus valores variando de -2 a +2, de acordo com a escala de Likert.

Verificando conceitos como usabilidade, conteúdo, layout e dificuldade, ditados no modelo que visa avaliar se um jogo: (a) consegue motivar os estudantes a utilizarem o material de aprendizagem; (b) proporciona uma boa experiência nos usuários; (c) se os alunos acham que estão aprendendo com o jogo[16].

Com a execução do questionário tem-se então os resultados obtidos através do mesmo e os dados retirados dessa analise.

1.7. Resultados

Identifica-se que 95,2% dos alunos achou que o conteúdo foi bastante reforçado, percebe-se que 64,2% achou o mesmo fácil de entender. Os dados também mostram que 67,8% dos alunos acharam que as informações passadas na fase não são confusas e 32,1% dos alunos achou a dificuldade da fase mediana.

Percebe-se que 42,8% dos alunos achou que o jogo é divertido, mas muito desafiador de acordo com 53,5% das respostas. Identifica-se que 60,7% dos alunos achou o layout, de forma geral, não dificultou o aprendizado do conteúdo e 31,1% dos alunos achou que as músicas do jogo estimulam em partes o ensino.

Com os dados analisados identifica-se que os desafios no ensino de redes tais quais: Dificuldade de motivar o aluno no aprendizado; Dificuldade de preparar o aluno para a prática profissional em um ambiente acadêmico e a quantidade de disciplinas que envolvem um número significativo de conceitos puramente técnicos, são tratados nesse trabalho e explorados em sua análise de dados.

1.8. Considerações Finais

Considera-se neste artigo que o desenvolvimento de um jogo educacional focado na disciplina de redes de computadores – mais especificamente, no funcionamento da arquitetura de comunicação TCP/IP – auxilia no processo de ensino/aprendizagem dos conteúdos de redes. O jogo é desafiador, explicativo e o conteúdo a referente com o visto na academia, além de garantir a interatividade e sociabilidade de um jogo educacional.

O jogo é Open Source, leve, usa a tecnologia mais adequada para o ensino da disciplina de redes de computadores, tem como proposta ensinar e reforçar o conteúdo da arquitetura TCP/IP nas disciplinas de redes e nos testes realizados em sala de aula obteve êxito no que se propôs a fazer.

Identifica-se também que os elementos problemáticos no ensino de redes, tais quais carência de material didático interativo, dificuldade de montagem de laboratório práticos e passividade do educando no ensino são retratados e levados em consideração durante o desenvolvimento, uso e avaliação do jogo.

Percebe-se ainda que uma grande porcentagem dos alunos viu no jogo uma forma fácil e prática de reforçar o conteúdo demonstrado em aula, uma porcentagem considerável dos acadêmicos achou o jogo divertido e desafiador.

Nota-se então que a demanda por novos materiais acadêmicos que ajudem os docentes/discentes, na área de redes de computadores é real e expressiva, agregando valor ao produto final e ao conhecimento.

Considera-se com este trabalho que o problema identificado, escassez de alternativas para o ensino de redes de computadores é abordado, e uma solução através de um jogo educacional é desenvolvida.

References

[1] Tanenbaum, Andrew. (2011). "S. Redes de Computadores. São Paulo: Ed.".

[2] Fall, Kevin R.; Stevens, W. Richard. (2011). “TCP/IP illustrated, volume 1: The protocols”. addison-Wesley,

[3] Antunes, C. (2011). “Jogos para a estimulação das múltiplas inteligências”. Editora

[4] Sá, E.J.V; Teixeira, J.S.F; Fernandes, C.T (2007) "Design de atividades de aprendizagem que usam Jogos como princípio para Cooperação", In: XVIII Simpósio Brasileiro de Informática na Educação (SBIE), São Paulo - SP, Brasil

[5] Torres, Gabriel. (2015). Redes de computadores. Novaterra Editora e Distribuidora LTDA.

[6] Gallo, Michael A. et al. (2003). “Comunicação entre computadores e tecnologias de rede”. Pioneira Thomson Learning.

[7] Santos, Walter. (2016). “Uso de simuladores como ferramenta no ensino aprendizagem de redes de computadores”. Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento, v. 4, n. 2.

[8] Sarkar, N. (2015). “Tools for Teaching Computer Networking And Hardware Concepts”, Information Science Publishing. Taiwan, 2006, v. 10, p. 268-270. Disponível em:<http://www.ifets.info/journals/10_1/24.pdf>. Acesso em: 10 junho

[9] Herpich, Fabrício, et al. (2013). "Jogos Sérios na Educação: Uma Abordagem para Ensino-Aprendizagem de Redes de Computadores (Fase I)." XVIII Conferência Internacional sobre Informática na Educação–TISE.

[10] Botelho, L. (2004). “Jogos educacionais aplicados ao e-learning”. Disponível em:http://www.elearningbrasil.com.br/news/artigos/artigo_48.asp. Acessado em março de 2016.

[11] Tarouco, Liane Margarida Rockenbach, et al. (2004). "Jogos educacionais." CINTED, UFRGS.

[12] Mattar, João. "Games em educação: como os nativos digitais aprendem." (2010).

[13] Andrade; Marlon Marcos; Igor de Oliveira Knop. (2015). "Projeto e desenvolvimento de jogos eletrônicos multiplataforma: um estudo de caso utilizando Cocos2d-x." Caderno de Estudos em Sistemas de Informação 2.1

[14] Alves, F. D. S. N. L. (2010). “Jogos Digitais e Aprendizagem: um estudo de caso sobre a influência do design de interface”. In Anais do SBGames 2010.

[15] Adams, E. (2009). “Fundamentals of Game Design”, New Riders, 2nd Edition.

[16] Irvine, Cynthia E., Michael F. Thompson, and Ken Allen. "CyberCIEGE: gaming for information assurance." IEEE Security & Privacy 3.3 (2005): 61-64.

[17] Jung, C. F. (2004). “Metodologia para pesquisa & desenvolvimento: aplicada a novas tecnologias, produtos e processos”. Rio de janeiro: Axcel Books, 2004.

[18] Aguiar, B., & Soares, N. (2012). Proposta de uma escala de autorrelato para a análise de jogos. Brasília: SBGames, 2012.

8. Ambientes de Programação em Blocos

Capítulo 8

Uma Análise de Ambientes de Programação em Blocos com Base em Recomendações de Interação Criança-Computador

Esteic Janaina Santos Batista[1], Anderson Corrêa de Lima[1]

[1] Universidade Federal de Mato Grosso do Sul, Ponta Porã, Brasil

Abstract

This work aims to perform an analysis of visual programming environments in blocks using Nielsen heuristics that aim to evaluate usability, and those of Malone that evaluate the fun. We also proposed a set of heuristics to evaluate interface design items for children based on the recommendations in Child-Computer Interaction. Environments were evaluated by specialists using the original and child heuristics, using simplified sentences, and the Fun-Sorter and Again-Again tools for pre-school children.

Resumo

Este trabalho tem como objetivo realizar uma análise de ambientes de programação visual em blocos utilizando heurísticas de Nielsen que visam avaliar usabilidade, e as de Malone que avaliam a diversão. Também foi proposto um conjunto de heurísticas para avaliar itens de design de interfaces para crianças com base nas recomendações em Interação Criança-Computador. Os ambientes foram avaliados por especialistas utilizando as heurísticas originais e por crianças, utilizando sentenças simplificadas, e as ferramentas Fun-Sorter e Again-Again para avaiação por crianças em idade pré-escolar.

1.1. Introdução

Iniciativas que fazem uso de computadores na escola, e mais recentemente de dispositivos móveis, óculos de realidade virtual e kits educacionais de robótica, têm sido apoiadas e impulsionadas por abordagens de desenvolvimento interacionistas [3].

Nos últimos dez anos, surgiram diversos ambientes de programação visual em blocos para crianças a partir de quatro anos de idade, eles têm como objetivo desenvolver o raciocínio lógico, pensamento computacional, colaboração, dentre outras habilidades.

Dada a recente multiplicidade de ambientes de ensino de programação para crianças, alguns questionamentos surgem, por exemplo: a) Como as crianças têm avaliado a usabilidade destes ambientes e o quão estão se divertindo? b) Quais elementos de design baseados em diretrizes de Interação Criança-Computador estão presentes nestes ambientes?

Neste trabalho, analisamos cinco ambientes de programação visual em blocos: Scratch, ScratchJr, AppInventor, Código do Flappy bird e a ferramenta LighBot, utilizando as heurísticas de Nielsen [10], que visam avaliar a usabilidade e as de Malone [7] que avaliam o divertimento. Propomos, também neste trabalho, um conjunto de heurísticas para avaliar itens de design para interfaces para crianças com base em recomendações de Interação Criança-Computador.

Os ambientes de programação visual em blocos foram avaliados por especialistas utilizando as heurísticas originais e por crianças por meio de frases simplificadas das heurísticas. Para que as crianças em fase pré-escolar pudessem participar da avaliação oram utilizadas as ferramentas Fun-Sorter e Again-Again.

1.2. Fundamentação Teórica

Esta seção apresenta o referencial teórico que norteia o desenvolvimento deste trabalho. Inicia-se apresentando as teorias educacionais que dão base para o uso de computadores e o ensino de programação nas escolas.

1.2.1. Construcionismo

Seymour Papert cunhou a teoria do Construcionismo, onde o indivíduo através do computador constrói seu próprio conhecimento de duas formas: quando o indivíduo cria um objeto no computador, como um jogo ou um programa, e quando ele cria algo do seu interesse e por isso torna-se mais motivado, pois para Papert, o envolvimento afetivo torna a aprendizagem mais efetiva [17] [11].

1.2.2. Ensino de Programação para Crianças

O ensino de programação para crianças e jovens já faz parte de diversas iniciativas realizadas no Brasil e em outros países, com a proposta de desenvolver o pensamento crítico, melhorar o raciocínio lógico, favorecendo assim a capacidade de busca de novas soluções frente aos problemas, com os quais o estudante venha a se deparar ao longo de sua vida. O conceito de Pensamento Computacional (PC) está intrinsecamente relacionado ao desenvolvimento destas habilidades [18].

1.2.3. Interação Criança-Computador

O campo da Interação Criança-Computador (ICC) é definido como uma subárea da Interação Homem-Computador, que diz respeito ao estudo da concepção, avaliação e implementação de sistemas computacionais interativos para criança e fenômenos que as rodeiam, anali-

sando o impacto mais amplo da ação da tecnologia nestas e na sociedade [6].

Muitas pesquisas de ICC tratam do estudo e desenvolvimento de tecnologias em que as crianças assumem um dos quatro papéis no processo de desenvolvimento de uma tecnologia levantado por Alisson Druin [4], sendo estes: parceira de design, informante, tester e usuário. Nesta pesquisa, elas assumiram papel de tester, ao avaliarem os ambientes de programação visual em blocos.

1.3. Metodologia

Esta seção apresenta os ambientes de programação que foram avaliados pelos especialistas e pelas crianças, e também as heurísticas e as ferramentas utilizadas na avaliação.

1.3.1. Ambientes de programação em blocos avaliados

Nesta pesquisa, quatro ambientes de programação visual em blocos voltados para o público infantil destinados à criação de jogos, animações, histórias e aplicativos; e um ambiente para resolução de desafios foram avaliados por crianças e especialistas. A seguir, uma breve descrição sobre os ambientes avaliados neste trabalho:

- Scratch 2: é um ambiente de programação visual em blocos para criação de jogos, animações e histórias, que deu base para diversas outras interfaces de programação. A ferramenta é destinada à crianças a partir de 8 anos [1] [14].
- ScratchJr: é uma ferramenta similar ao Scratch, destinado à crianças a partir de 4 anos. A ferramenta não possui texto nos seus comandos, o que facilita a utilização por crianças, que ainda não possuem habilidade de leitura [2].
- AppInventor 2: ambiente de programação visual em blocos destinado ao desenvolvimento de aplicações para Android.
- Código do Flappy bird do Code.Org: o Código do Flappy bird é um dos cursos de Code.org, projeto que tem a finalidade de ensinar programação para pessoas de qualquer idade. No curso do Flappy Bird há 10 atividades a serem realizadas, sendo as primeiras destinadas ao aprendizado de comandos para personalizar o jogo do Flappy Bird.
- LightBot: é um aplicativo para dispositivos móveis que tem diversos desafios em que o usuário deverá resolver manipulando um robô. Para manipular o robô são disponibilizados comandos de programação.

1.3.2. Avaliação com Heurísticas

A avaliação de interfaces utilizando heurísticas é realizada geralmente por um grupo de especialistas que analisam uma interface considerando um conjunto de heurísticas adequadas ao domínio e contexto da aplicação [16]. A avaliação heurística é feita examinando uma interface e tentando chegar a uma opinião sobre o que é bom e ruim nela.

As ferramentas Scratch, ScratchJr, AppInventor e Flappy bird do Code.org foram avaliadas por quatro acadêmicos dos cursos de Sistemas de Informação, Ciência da

Computação e Pedagogia, que já tinham experiência com avaliação de softwares. Estes assumiram o papel de especialistas na avaliação. Nesta avaliação foram utilizados três conjuntos de heurísticas originais:

1.3.2.1. Heurísticas de Nielsen

Jakob Nielsen propôs uma lista de 10 heurísticas para serem utilizadas no design de interação [9] [10] e avaliação de uma interface. Desde então, diversos trabalhos as referenciam e utilizam na avaliação de softwares.

1.3.2.2. Heurísticas de Malone

Malone formulou um conjunto de heurísticas que permitem avaliar o divertimento proporcionado por uma interface com o usuário [7]. As heurísticas são separadas em três categorias sendo elas desafio, fantasia e curiosidade.

1.3.2.3. Heurísticas de itens de design baseadas na Interação Criança-Computador

enquanto as diretrizes de Nielsen e Malone avaliam a usabilidade e diversão, propomos neste trabalho heurísticas baseadas em diretrizes de Interação Humano-Computador, que foram fundamentas em trabalhos centrais de avaliação de softwares infantis e pesquisas relacionadas [5] [15] [13] [8] sendo o livro de Interação Criança-Computador o principal deles, que foi traduzido do inglês para o português brasileiro para dar apoio aos fundamentos destes itens e do estudo em geral[6]. Os elementos de design selecionados que compõem as heurísticas são: *affordance*, *feedback*, navegação, gestos, sons, cores, tipografia, ícones personagens, participação dos pais e suporte a colaboração e comunicação.

As Heurísticas de avaliação de itens de design com base na Interação Criança-Computador foram adaptadas para as crianças por meio de frases simplificadas correspondentes a cada heurística. Para a avaliação de heurísticas com crianças no papel de especialistas foram escolhidas escolas que tinham projetos extraclasse de programação, por já conhecerem ferramentas de programação em blocos. Esta avaliação envolveu 20 alunos do Ensino Fundamental 2 a partir de 10 anos de idade de duas escolas públicas e uma particular.

1.3.3. Avaliação com Fun-Sorter e Again-Again

Os softwares ScratchJr e LightBot foram avaliados por crianças do Jardim II, que tem em média 5 anos de idade. No total, 12 crianças avaliaram os ambientes utilizando o método FunSorter, e 8 crianças avaliaram utilizando o método Again-Again. Estes dois métodos foram propostos no trabalho de Janet Read [12].

O Fun-Sorter consiste na avaliação comparativas entre tecnologias conectadas ou concorrentes. Esta ferramenta pode ajudar o avaliador a determinar quais funcionalidades e/ou ferramentas no software podem ser mais atraentes ou menos divertidas, por exemplo.

No Again-Again, a criança marca sim, talvez ou não, para cada atividade apresentada de uma tecnologia específica, levando em consideração para cada uma delas a pergunta ”Você gostaria de fazer isso de novo?”. O motivo para isso é que, para a maioria das crianças, uma atividade divertida seria desejada ser repetida [12].

Na avaliação do Scratch e Lightbot com o FunSorter por crianças pré-escolares, foram fornecidos cartões com telas da interface de ambas as ferramentas. Elas deveriam dispor os cartões da interface na tabela do FunSorter conforme suas opiniões.

No Again-Again foram fornecidas frases referentes a atividades apenas do ScratchJr. As crianças deveriam preencher a Tabela indicando se gostariam de realizar a atividade novamente.

1.4. Análise dos Resultados

Esta seção apresentará o resultado das avaliações das ferramentas realizadas por especialistas e pelas crianças utilizando as heurísticas simplificados e com as ferramentas FunSorter e Again-Again.

1.4.1. Avaliação com heurísticas

Ao analisar as heurísticas de cada conjunto da tabela apresentada na Figura **??**, podemos verificar que nas heurísticas de Nielsen a heurística 2 foi a que foi mais relacionada nos problemas levantados que diz respeito à linguagem do usuário. Isto deve-se ao fato dos softwares não estar na língua materna das crianças do contexto da pesquisa realizada. Em relação a usabilidade, outra heurística diz respeito à criança conseguir reverter uma ação.

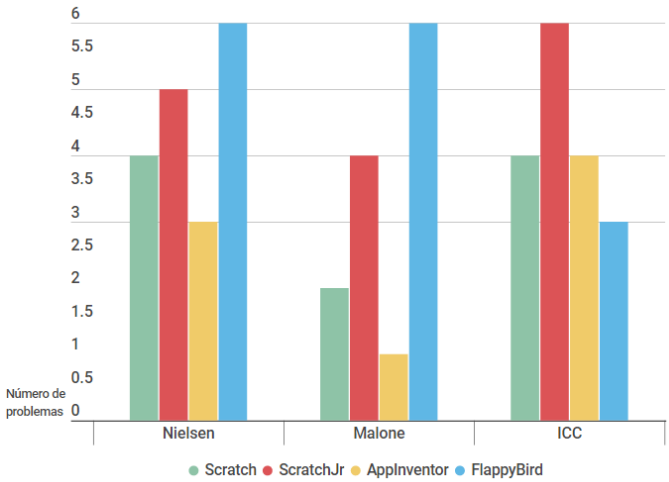

Figura 1.1. O gráfico apresenta o número de problemas encontrados por ambiente de programação para cada conjunto de heurísticas na avaliação realizada por especialistas.

Nas heurísticas simplificadas, utilizadas nas avaliações pelas crianças, Ao observamos as heurísticas que foram mais relacionadas aos problemas encontrados para cada conjunto de heurísticas apresentado no gráfico da Figura 1.2 podemos constatar que o conjunto de heurísticas de itens de design de ICC foram os que as crianças conseguiram

relacionar a maior parte dos problemas, e pouco relacionou-os às heurísticas de diversão de Malone. Desta forma, podemos sugerir que as crianças se divertem ao utilizar os ambientes de programação visual em blocos, apesar das dificuldades encontradas em relação aos itens de design da ferramenta.

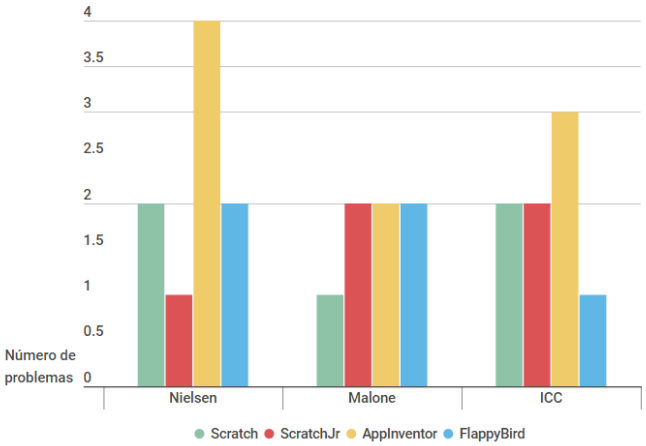

Figura 1.2. O gráfico apresenta o número de problemas encontrados por ambiente de programação para cada conjunto de heurísticas na avaliação realizada pelas crianças.

1.4.2. Avaliação com a ferramenta Fun-Sorter e Again-Again

Na avaliação realizada com as doze crianças pré-escolares utilizando o método Fun-Sorter observou-se conforme apresentado no gráfico da Figura 1.3 que as crianças consideraram as Telas 2 e 3, que correspondem às telas de escolha de um personagem no ScratchJr e uma atividade das instruções básicas, com as atividades mais difíceis e divertidas. Desta forma pode ser interessante atividades que desafiem as crianças.

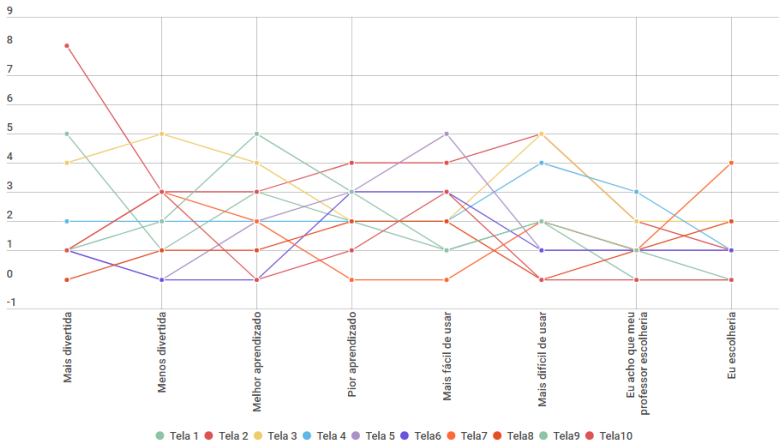

Figura 1.3. Resultado da avaliação realizada com crianças do Jardim II utilizando a ferramenta Fun-Sorter.

Na avaliação de atividades do ScratchJr com a ferramenta Again-Again podemos

verificar algumas crianças apontaram que não gostariam de realizar a ação de excluir o bloco, que pode estar relacionado com o problema que os especialistas apontaram de que a forma de exclusão do bloco não é intuitiva e nem consistente na ferramenta. Em relação à atividade de inserir uma nova página, poucas crianças gostariam de realizar novamente. Durante a observação da utilização da ferramenta pode-se perceber que elas ficavam confusas quanto a este recurso, pois quando é inserida uma nova página, a criança deve incluir novos atores nela, desta forma, não há relação entre o ator de uma página com outra, fazendo com que a criança tenha que para cada nova página programar os atores, mesmo quando o objetivo é apenas alterar o fundo em determinado momento.

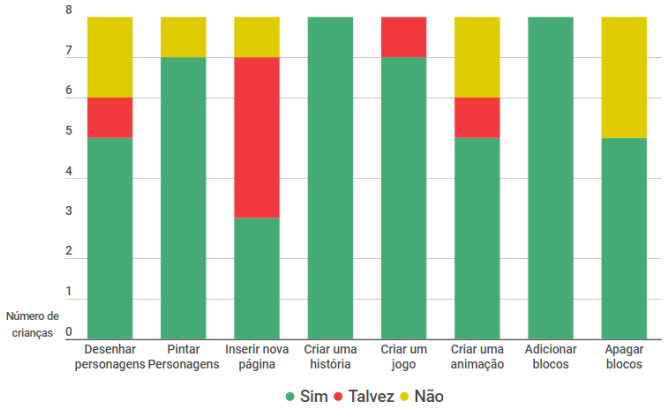

Figura 1.4. Resultado da avaliação realizada com crianças do Jardim II utilizando a ferramenta Again-Again.

1.5. Considerações Finais

Os ambientes de programação visual em blocos foram avaliados neste estudo por crianças e especialistas utilizando heurísticas, nas quais foram utilizadas as heurísticas de Nielsen para avaliar a usabilidade e as heurísticas de Malone para avaliar a diversão. Além destas, foram propostas neste trabalho onze heurísticas para avaliar itens de design de interfaces infantis com base nas recomendações de ICC, que foram capazes de apontar diversos problemas, obtendo desta forma resultados satisfatórios para a primeira utilização. Em trabalhos futuros, pretende-se reaplicar elas novamente com um grupo maior e mais heterogêneo.

Os métodos de avaliação Fun-Sorter e Again-Again para a avaliação de dois ambientes voltados para crianças pré-escolares permitiram encontrar alguns problemas de usabilidade e itens de design, e poucos problemas em relação a diversão.

Durante a avaliação realizada neste trabalho observou-se a ausência de recursos nos ambientes de programação visual que permitam a participação dos pais e a colaboração de uma criança com outros colegas distantes fisicamente e até mesmo quando estes estão próximos. Desta forma, torna-se necessário que os ambientes de programação visual em blocos sejam desenvolvidos para incorporar estes recursos deste o início do desenvolvimento do ambiente ou ferramenta de programação para crianças.

Referências

[1] Esteic Janaina Santos Batista, Amaury Castro Jr, Andreia Alfonso Larrea, and Cintia Adriana Canteiro Bogarim. Utilizando o scratch como ferramenta de apoio para desenvolver o raciocínio lógico das crianças do ensino básico de uma forma multidisciplinar. In *Anais do Workshop de Informática na Escola*, volume 21, page 350, 2015.

[2] Marina Umaschi Bers and Mitchel Resnick. *The official ScratchJr book: Help your kids learn to code*. No Starch Press, 2015.

[3] Cláudia Davis and Zilma de Oliveira. *Psicologia na educação*. Cortez, 1994.

[4] Allison Druin. The role of children in the design of new technology. *Behaviour and information technology*, 21(1):1–25, 2002.

[5] Lucas Aviles Fabossi and ALSV Guimarães. Design de interface voltado a crianças em educação infantil. *Universidade Tecnológica Federal do Paraná*, 2014.

[6] Juan Pablo Hourcade. Child-computer interaction. *Self*, 2015.

[7] Thomas W Malone. Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 conference on Human factors in computing systems*, pages 63–68. ACM, 1982.

[8] A Mano. Interfaces de computador para crianças–avaliação e construção. *Universidade do Minho*, 2005.

[9] Jakob Nielsen. *10 Usability Heuristics for User Interface Design*, 1995 (accessed November 14, 2017). `https://www.nngroup.com/articles/ten-usability-heuristics/`.

[10] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM, 1990.

[11] Seymour Papert. A máquina das crianças. *Porto Alegre: Artmed*, 1994.

[12] Janet C Read. Is what you see what you get? children, technology and the fun toolkit. In *International Workshop on*, page 67, 2008.

[13] Cátia Andreia Tavares dos Santos Resende. Design de interacção centrado nas crianças: estudo do caso biblon. Master's thesis, Universidade de Aveiro, 2010.

[14] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, et al. Scratch: programming for all. *Communications of the ACM*, 52(11):60–67, 2009.

[15] Sesame Street. Best practices: Designing touch tablet experiences for preschoolers. In *Domestic Research Sesame Workshop http://www. sesameworkshop. org/wp_install/wp-content/uploads/2013/04/Best-Practices-Document-11-26-12. pdf*, 2012.

[16] João Filipe Pereira Valente. Avaliação da usabilidade e diversão em interfaces web para criançascaso de estudo escolinhas. pt. *Mestrado Integrado em Engenharia Informática e Computação Tese de Mestrado Integrado em Engenharia Informática e Computação), Universidade do Porto, Faculdade de Engenharia*, 2011.

[17] José Armando Valente et al. Informática na educação: instrucionismo x construcionismo. *Manuscrito não publicado, NIED: UNICAMP*, 1997.

[18] Jeannette M Wing. Computational thinking. *Communications of the ACM*, 49(3):33–35, 2006.

9. Reconhecimento de Sinais de Transito

Chapter 9

Reconhecimento de Sinais de Trânsito Utilizando *Deep Learning*

Rafael Ribeiro da Silva e Orlando Bisacchi Coelho

Abstract

This paper presents a solution based on Deep Learning for the Automatic Traffic Sign Recognition task using a Convolutional Neural Network. The experiments reported herein, developed in Keras and TensorFlow, performed a comprehensive search for hiperparameters values for the ADAM and SGD optimisation algorithms. The best result achieved an accuracy of 97,02%. The model recognised a total of 12.254 signs in a benchmark dataset containing 12.630 signs. Such a result is similar to the results produced in The German Traffic Sign Recognition Benchmark, a multi-category classification competition held at IJCNN 2011.

Resumo

Este artigo apresenta uma solução baseada em Deep Learning para a tarefa de Reconhecimento Automático de Sinais de Tráfego usando uma Rede Neural Convolucional. O experimento aqui reportado, desenvolvido em Keras e Tensorflow, realizou uma busca sistemática por valores de hiperparâmetros para os algoritmos de otimização ADAM e SGD. O melhor resultado conseguiu uma acurácia de 97,02%. O modelo reconheceu um total de 12.254 sinais em um dataset de referência contendo 12.630 sinais. Tal resultado é semelhante aos resultados produzidos em The German Traffic Sign Recognition Benchmark, uma competição de classificação multicategorias realizada no IJCNN 2011.

9.1. Introdução

Por ter um impacto direto com a manutenção da mobilidade, redução de custos causados pelos congestionamentos, redução de emissão de combustíveis e diminuição das mortes

no trânsito, entre outros pontos [Anderson et al. 2016], o estudo e desenvolvimento de carros autônomos tem crescido bastante nos últimos anos. Esses carros são compostos por vários sistemas, que desempenham todas as tarefas que são realizadas por um motorista em um carro comum. Uma dessas tarefas – o foco deste artigo – é o reconhecimento automático de sinais de trânsito. Essa é uma tarefa essencial a ser realizada pelos carros autônomos.

Este artigo apresenta uma solução computacional para o problema do reconhecimento de sinais de trânsito baseado em *deep learning* [LeCun, Bengio and Hinton 2015]. O sistema é capaz de reconhecer uma porcentagem extremamente grande dos sinais. É importante observar que este trabalho considera já realizadas a detecção e a normalização em tamanho e orientação dos sinais de trânsito.

A primeira seção deste artigo descreve o referencial teórico que guiou o trabalho. Em seguida, os aspectos metodológicos, tais como a arquitetura computacional escolhida e o espaço de hiperparâmetros que foi explorado, são descritos. Em sequência, são apresentados os resultados obtidos a partir dos vários testes feitos. Por último, é feita uma discussão dos resultados obtidos e são apresentadas sugestões para futura continuação do trabalho.

9.2. Fundamentação Teórica

Como o reconhecimento foi feito a partir de imagens, a utilização de técnicas de *Deep Learning* é muito recomendada [LeCun, Bengio and Hinton 2015]. Segundo LeCun e colegas [2015], Redes Neurais Convolucionais são a abordagem computacional mais bem sucedida na tarefa de reconhecimento de imagens. De fato, Redes Convolucionais estão por trás de uma série de avanços em reconhecimento de imagens, tais como a classificação de imagens em bancos de imagens com milhões de imagens pertencentes a milhares de classes distintas [LeCun et al. 1998].

9.3. Metodologia

9.3.1. Conjunto de Dados

Redes Neurais Convolucionais são sistemas de aprendizagem supervisionada: elas aprendem a partir de exemplos de entrada e seus respectivos rótulos. Torna-se então necessário a utilização de um conjunto de dados rotulados para ser utilizado no processo de treinamento e teste da rede. Foi utilizado neste artigo o GTSRB [Stallkamp et al. 2011] contendo sinais de trânsito alemães; este é o melhor conjunto de treinamento disponível publicamente para esta aplicação. (Vale notar, a propósito, que infelizmente não existe ainda um conjunto de treinamento disponível baseado nos sinais de trânsito usados no Brasil.) O conjunto de dados contém 43 placas de trânsito. Esse conjunto de dados possui dois subconjuntos, um de treinamento e outro de teste que possuem respectivamente 39.209 e 12.630 imagens de tamanhos variados. Devido a essa variação, as imagens precisaram ser tratadas para que todas ficassem com o mesmo tamanho: 40x40 pixels. A Fig. 9.1 mostra alguns exemplos do *dataset* utilizado.

Figura 9.1. Exemplos de imagens no *dataset* utilizado.

9.3.2. Configuração da Rede

O modelo da rede foi baseado na LeNET-5, desenvolvida por Yann LeCun [Krizhevsky, Sutskever and Hinton 2012] para solução de problema de reconhecimento de dígitos. A Fig. 9.2 mostra a arquitetura da rede. A imagem é representada por três matrizes de 40 x 40 pixels, cada uma correspondendo a um canal de cores. A primeira camada de processamento é composta por 32 mapas que extraem características da imagem, usando para tal a operação de convolução, com base em núcleos de convolução de tamanho 5 x 5 pixels. A próxima camada é composta por 32 mapas que extraem outras características da imagem, desta vez usando a operação *max-pooling* sobre matrizes de 2 x 2 pixels. Seguem-se mais duas camadas, realizando convolução e *max-pooling*, com número de mapas e dimensões tal como descritas na figura. Por fim, uma camada totalmente conectada de 1.000 unidades projeta em uma camada de saída com 43 unidades. A rede toda foi treinada por *backpropagation* [Rumelhart, Hinton and Williams 1986].

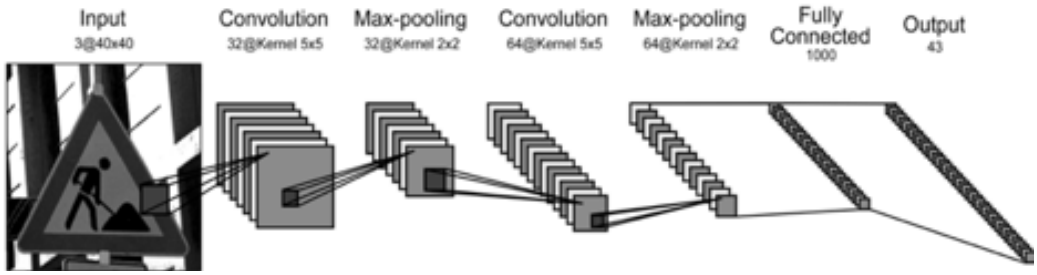

Figura 9.2. Modelo da Rede, que recebe uma imagem 40x40 e retorna as probabilidades de pertencimento da imagem a cada uma das 43 classes.

9.3.3. Ambiente de Desenvolvimento

Para a criação, treinamento e teste da rede da rede foi utilizada a Interface de Programação de Aplicações de *deep learning* Keras [Chollet et al. 2015], utilizando-se o ambiente para desenvolvimento de *deep learning* TensorFlow [Abadi et al. 2016] como *backend*, bem como a plataforma Anaconda, que gerencia os pacotes necessários para a execução do código, escrito em Python. Para executar os testes foi utilizado um MacBook Pro com macOS 10.13.3.

9.3.4. Algoritmos de Aprendizagem e Hiperparâmetros

Para o treinamento da rede foram utilizadas variações do algoritmo *backpropagation* implementadas nos algoritmos ADAM [Kingma and Ba 2014] e SGD [Ruder 2010], que, segundo os testes feitos por Kingma e Ba [2014] foram os que obtiveram melhores resultados.

Após a escolha do algoritmo de aprendizagem, é necessário escolher a melhor configuração para o algoritmo. Isso se dá encontrando as melhores combinações dos hiperparâmetros do algoritmo em questão.

Os valores *default* sugeridos para os hiperparâmetros do algoritmo ADAM são: *Learning Rate* = 0,001; Beta 1 = 0,9 e Beta 2 = 0,999. Já os valores *default* para o algoritmo SGD são: *Learning Rate* = 0,01; *Momentum* = 0 e Nesterov Inativo.

Decidiu-se por experimentar com pequenas variações em torno dos valores sugeridos, com o objetivo de testar o máximo de combinações possíveis, a fim de achar a melhor combinação de hiperparâmetros. *Learning Rate* foi variada entre 0,001 e 0,4. Para ADAM, os valores testados de Beta 1 e Beta 2 variaram entre 0,9 e 0,99955. Para o algoritmo SGD o valor de *Momentum* foi variado entre 0,1 e 0,99. O parâmetro Nesterov foi testado tanto como Ativo como Inativo. Explorar essas combinações de hiperparâmetros necessitou de um total de mais de 350 testes.

9.4. Resultados

O melhor resultado nos experimentos, em termos de acurácia, foi obtido com o algoritmo SGD: 97,02%; dos 12.630 exemplos no *dataset*, a rede classificou corretamente 12.254. Tais resultados foram obtidos com os valores dos hiperparâmetros descritos na Tabela 9.1.

Tabela 9.1. Melhor solução encontrada: algoritmo de aprendizagem, valores dos hiperparâmetros e acurácia.

Algoritmo	Learning Rate	Momentum	Nesterov	Acurácia
SGD	0,03	0,85	Ativo	97,02

A matriz de confusão mostra se uma determinada entrada está sendo confundida com uma outra específica. Isso pode acontecer se as imagens utilizadas no treinamento não são boas o suficiente, fazendo com que a rede não consiga distinguir entre imagens distintas. Mais importante, pode apontar os exemplos que a rede está tendo mais dificuldade de aprender. A Fig. 9.3 mostra a matriz de confusão gerada para o teste da rede com a configuração que obteve o melhor resultado, utilizando todo o conjunto de teste.

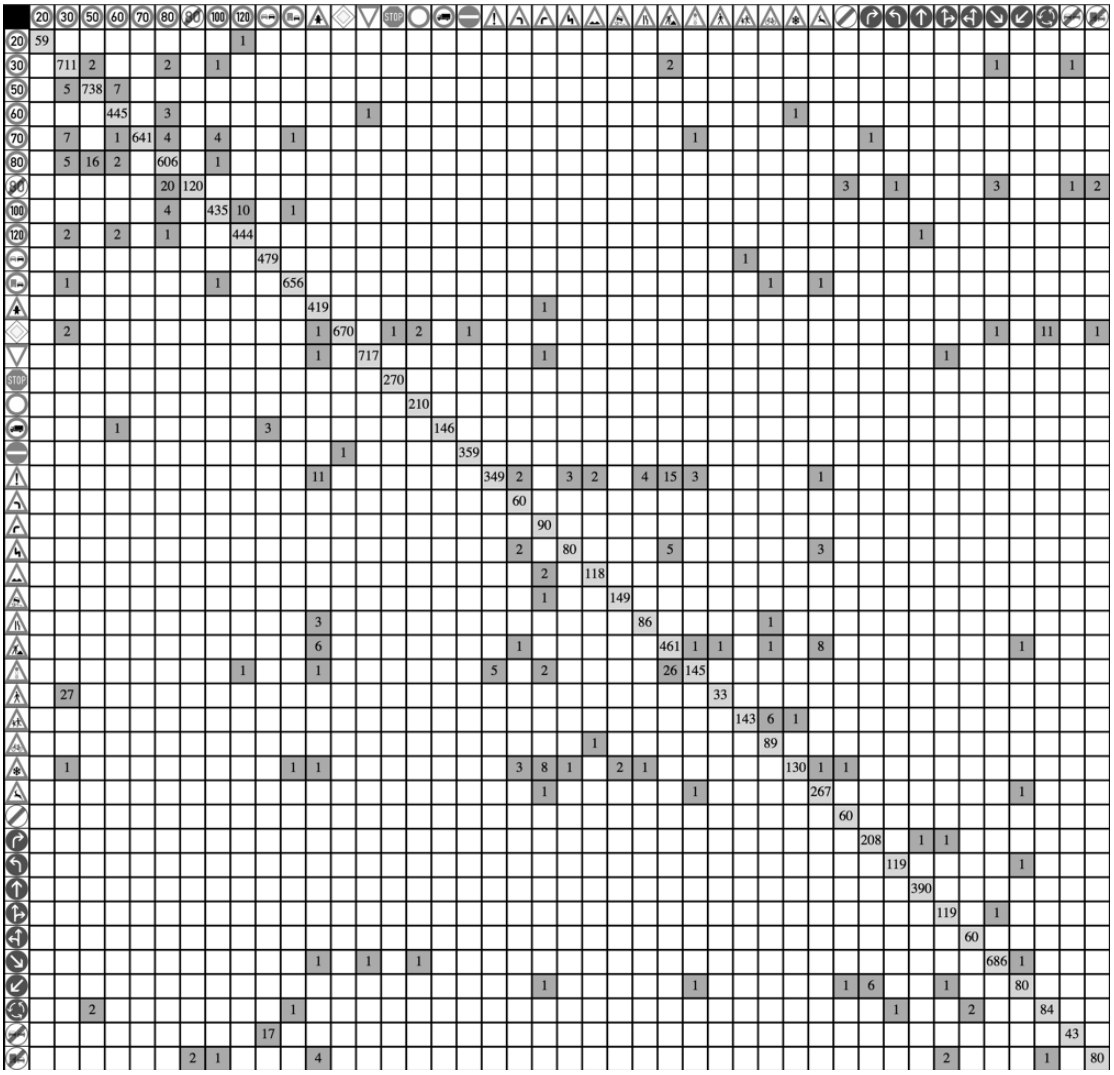

Figura 9.3. Matriz de confusão para o melhor resultado obtido.

A rede teve maior dificuldade com os sinais dos tipos proibitivos, especialmente de velocidade, e com os sinais de perigo (triangulares), devido à semelhança existente entre os sinais de cada classe. Fatores que podem ter influenciado para a ocorrência desses erros são a baixa qualidade das imagens de entrada e, eventualmente, a condição de iluminação da placa quando da obtenção da imagem.

9.5. Discussão

Apesar da relativa facilidade de se conseguir resultados satisfatórios, não há uma configuração específica que resolva qualquer problema. Deste modo, se uma configuração é boa para um determinado problema, pode ser que essa mesma configuração não seja adequada para outro problema distinto. Com isso, para garantir o melhor resultado possível, é necessário variar e testar o máximo possível de configurações. Só assim é possível aumentar o grau de certeza de que o melhor resultado possível foi obtido.

Tão importante quanto a configuração é o *dataset* utilizado, que, se não for grande o suficiente e de boa qualidade, vai impactar no desempenho do treinamento da rede. É importante observar que, no momento, dentro do conhecimento dos autores, não existe um *dataset* publicamente disponível melhor do que o utilizado neste trabalho.

O resultado final obtido ainda não é ideal para se utilizar em produção, porém é um valor bem aceitável para um estudo inicial. A acurácia alcançada neste trabalho, de 97,02%, foi bem próxima ao melhor resultado obtido no *workshop* para o qual foi criado o *dataset* aqui utilizado [Stallkamp et al. 2011], no qual os resultados obtidos variaram entre 73,89% e 98,98%, com mediana de 97,63%. O resultado aqui reportado não está muito longe da performance humana – estimada entre 97,39% e 99,90, dependendo do tipo de sinal [Stallkamp et al. 2011] – nem tão longe do estado da arte – 99,94% [Shustanov and Yakimov 2017].

Resultados superiores ao obtido neste trabalho poderiam ser obtidos a partir de dois melhoramentos: um *dataset* com imagens com maior resolução (preferivelmente incluindo os sinais de trânsito usados no Brasil) e variações da arquitetura computacional usada, em termos de número de camadas e número de unidades por camada – o que só pode ser investigado experimentalmente.

9.6. Conclusão

É muito simples começar a trabalhar com Redes Neurais Convolucionais. Além do mais, mesmo experimentos relevantes podem ser desenvolvidos sem que se possua um equipamento potente. Resultados relativamente próximos ao estado da arte – que já excede a performance humana – podem ser obtidos, se o treinamento da rede é feito com a devida atenção.

Uma sugestão para a continuidade deste trabalho consiste no desenvolvimento de um novo *dataset*, contendo os sinais de trânsito oficiais usados no Brasil, e o desenvolvimento de uma solução semelhante para esse *dataset*.

O desenvolvimento de abordagens mais robustas do que a aqui proposta – ou até mesmo que avancem o estado-da-arte – demandará a existência de um *dataset* de referência que inclua placas obtidas nas condições atmosféricas mais variadas (por exemplo, com neblina ou chuva forte) e que critérios de processamento em tempo real sejam levados em consideração.

9.7. Referências

Abadi, M. et al. (2016) “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”, unpublished, arxiv.org/pdf/1603.04467.pdf.

Anderson, J. M., Kalra, N., Stanley, K., Sorensen, P., Samaras, C. and Oluwatola, T. A. (2016) “Autonomous vehicle technology: a guide for policymakers”, Rand Corporation.

Chollet, F. et al. (2015) “Keras”, unplublished, https://keras.io.

Kingma, D. P. and Ba, J. L. (2014) “Adam: a method for stochastic optimization. Proc. 3rd. Intl. Conf. Learning Representations (ICLR).

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) “Imagenet classification with deep convolutional neural networks”, Proc. Advances in Neural Information Processing Systems 25 (NIPS 2012).

LeCun, Y., Bengio, Y. and Hinton, G. E. (2015) “Deep learning”, Nature, 521(7553), pp. 436-444.

LeCun, Y., Bottou, I., Bengio, Y. and Haffner, P. (1998) “Gradient-based learning applied to document recognition”, Proc. IEEE, vol. 86, no. 11, pp. 2278–2324.

Ruder, S. (2010) “An overview of gradient descent optimization algorithm”. arXiv:1609.04747v2, unpublished.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) “Learning representations by error back-propagation”, Nature, vol. 323, pp. 533-536.

Shustanov, A. and Yakimov, P. (2017) “CNN Design for Real-Time Traffic Sign Recognition”, 3rd. Intl. Conf. Information Technology and Nanotechnology.

Stallkamp, J. Schlipsing, M., Salmen, J. and Igel, C. (2011) “The German traffic sign recognition benchmark: a multi-class classification competition. Proc. Intl. Joint Conf. Neural Networks 2011 (IJCNN 2011), pp. 1453-1460.

10. Classifying Social Network Messages

Chapter 10

Classifying Social Network Messages to Support Emergency Responders in Natural Disasters

Bruno Sena, Elisa Yumi Nakagawa

Abstract

The amount of natural disasters in the last years has been increasing, requiring efficient mechanisms to detect and provide rapid responses to assist victims. In this scenario, social networks have been playing an important role by providing real-time information related to the such disasters. Due to the amount of data which are generated by these networks, it is important to select and prioritize relevant information. Thus, there is a need of software systems that can handle such information and support emergency responders (e.g., firefighters and hospitals) in disasters situations. In this scenario, several systems to aid those emergency services during critical situations were proposed, detecting messages related to such disasters; however, they do not indicate regions where assistance should be a priority. Therefore, in this project we developed a service that use social network messages to detect emergency situations and classify such messages into two different impact levels (i.e., low and high impact), thus indicating priority regions for the emergency responders. We evaluated our service using data collected from twitter related to the Hurricane Irma, which hit Miami region in September 2017. As results, we identified that our service can effectively detect the hurricane occurrence with precision of 76%, and classify the observation and high impact messages with precision of 55% and 67%, respectively.

1.1. Introduction

Natural disasters, such as earthquakes, floods, hurricanes and tsunamis have been causing massive damage and threaten millions of people lives and infrastructure for several years [Wex et al. 2014]. The occurrence of these events are often accompanied by the uncertainty of how intense they will be and the possibility of serious losses if not responded to it properly [Yang et al. 2013]. To minimize damages and injuries caused by such disasters, and also to save people's lives, researchers have focused attention on the development of emergency response systems, which support decision-making and the management of crisis situation [Yang et al. 2013].

Disasters management systems are developed, used and improved by organizations to assist in the response to an emergency situation, supporting communications, data gathering, data analysis, and decision making [Phillips 2005]. Although rarely used, such systems must be accurate and efficient. Moreover, because they deal with people's lives in a critical domain, such systems cannot fail. These systems are used to automatically extract structured knowledge from unstructured information, and detect when response actions are required. For this, collaboration between numerous people and groups is needed [Kyng et al. 2006].

In this context, social networks are playing an important role in people's social lives nowadays due to the rapid growth of the Internet technology in recent years [Zareie and Sheikhahmadi 2017]. People usually use social networks to share information about their lives, photos, videos, and opinions. However, due to the rapid propagation of such information, these networks have become an important way to share information on the web, especially during critical events such as natural disasters (i.e., hurricanes, earthquakes, flood) [Imran et al. 2013a], fact that made the task of turning the social media data into useful knowledge to emergency responders a strong focus of research [Zareie and Sheikhahmadi 2017].

In crisis situation, people turn to social media for exchanging information about the incident, or to express their views about, for instance, relief works and the government role in the response and recovery efforts [Palen et al. 2010]. Messages exchanged through the social media contain, sometimes, valuable information that can contribute to a situational awareness during an incident if analyzed in a timely and efficient manner [Starbird et al. 2010, Latonero and Shklovski 2010, Chowdhury et al. 2013]. Twitter has become a major channel for communication during natural disasters. Users have been using it to spread news about casualties and damages, donation efforts and alerts [Ashktorab et al. 2014, Imran et al. 2013b].

In this scenario, emergency response systems can use social network data to help in the identification of regions where there are people who need assistance and then issuing alerts. However, during such events the number of posts increase and it is difficult for the stakeholders (i.e., police, firefighters, doctors) to mine useful information. For this reason, alerts must be classified considering their impact level and then support the stakeholders in prioritizing the assistance, and, consequently, improving the quality of the service. In this sense, most emergency response systems developed can effectively detect or predict the occurrence of natural disasters, but do not provide information about regions where assistance should be a priority.

Therefore, the main objective of this project was to establish a means by which messages collected from social networks could be classified in different impact levels, thus indicating the priority regions for emergency responders. For this, we developed a service that collect messages from twitter for detecting the occurrence of a natural disaster, and classifying such messages into two different classes: (i) low impact: messages that present information but do not indicate any risk to people's lives; and (ii) high impact: messages that indicate situations of high risk for people's lives and infrastructures. To evaluate the service, we used data collected from twitter related to the Hurricane Irma, which hit Miami region in September 2017.

The remaining of this paper is organized as follows. Section 2 presents the related works. Section 3 presents the research steps to develop the service. Section 4 presents the results of our evaluation, and some discussions. Finally, in Section 5 we present the main conclusions.

1.2. Related Works

Several works in literature have addressed the power of analyzing social networks on the dissemination of information [Kwak et al. 2010]. This information source has been the focus of researchers and practitioners to support emergency managers and responders in the detection and analysis of natural disasters [Hui et al. 2012].

Intelligence techniques such as machine learning and statistics have made great advances in processing and classifying messages generated by social networks in crisis situations [Neppalli et al. 2017]. Sakaki et al. (2010) used these techniques to detect earthquakes in Japan through twitter data analysis. Moreover, some works used it to study the propagation of rumors during the Chilean earthquake [Mendoza et al. 2010] while others focused on analyzing the information credibility, developing automatic methods to assess it by using features extracted from user's behavior and tweets' social context [Castillo et al. 2011]. Ashktorab et al. (2014) combined classification and clustering techniques to acquire actionable information for disasters responders at runtime, and Neppalli et al. (2017) used such techniques to perform a geographical sentiment analysis of tweets during Hurricane Sandy.

Regarding the classification of short messages collected from twitter, Li et al. (2015) used a domain analysis to study the usefulness of labeled data from natural disasters, together with unlabeled data from a target disaster to learn classifiers for the target and concluded that disasters historical information can be useful for classifying new target problems. Besides that, Delany et al. (2005) and Cornack et al. (2007) compared machine learning techniques (Naive Bayes, SVM, logistic regression, and decision trees) to identify spam messages, considering several feature representations such as punctuations, upper case letters, bi-grams, and so forth.

Caragea et al. (2011) used such techniques to build models for the classification of messages from the Haiti earthquake into classes which represent people's most urgent needs. Ten classes were considered: i) medical emergency; ii) people trapped; iii) food shortage; iv) water shortage; v) water sanitation; vi) shelter needed; vii) collapse structured; viii) food distribution; ix) hospital services; and x) person news. They obtained a overall accuracy of 47% classifying messages by keywords, and 59% using feature abstraction.

Our approach differs from the aforementioned works because it presents a service not only for detecting the occurrence of a disaster, but also by classifying the messages using machine learning techniques into specific classes that can support emergency responders during natural disasters, in the identification of priority regions for assistance.

1.3. Research Steps

In this section we present the followed steps to develop the service, namely, (i) data collection; (ii) features reduction; (iii) model creation; (iv) machine learning technique selection; and (v) implementation.

1.3.1. Data Collection

Considering the impact that Irma had recently in the Caribbean sea, data related to such hurricane were considered as basis for the system. To get such data, we used the Tweepy[1] API, a complete set of twitter libraries for Python. Twitter was considered as the main data source due to the fact that is the most used social network during disasters [Li et al. 2015]. We investigated the hurricane path to define the window that defines the most accurate region to be used as an input for data collection. As result, we collected a set of 124874 tweets[2] from September 07 to September 15, the days when the hurricane hit the selected region.

1.3.2. Features Reduction

Due to the fact that textual processing techniques are very sensitive to the occurrence of terms in text classification, a pre-processing task is required. Analyzing data that have not been processed can produce misleading results. Thus, the following tasks were performed to pre-process the collected messages.

Stop Words: Stop words are the extremely common words in a vocabulary (e.g., and, so, some, such) which have little value in helping to select documents, analyzing contexts or in building a classifier. These words do not bring benefits to the classification, but decrease the performance of the classifier, therefore they are removed.

Message elements: Some message elements such as hyperlinks, emoticons and punctuation were removed. Hyperlinks usually contain information about the context in which the message is inserted, and in some messages the text does not make sense without the reference in the hyperlink. However, to analyze messages and use such information in classification is a challenge not addressed in this study. Punctuations are also removed due to the fact that if spaced from the text, the classifier could consider them as words. Emoticons only should be considered as a feature in a sentiment analysis, which is not the objective of this study.

1.3.3. Model Creation

Through domain analysis, and supported by experts opinions, we identified two potential impact levels in which the classification could be performed to support emergency responders:

Low impact: it is based on information and does not present any risk to people's lives or damage in infrastructures. Such alerts should be issued during disaster preparedness, storage of resources, i.e., water and food, and identification of shelters for protection;

High impact: messages that identify occasional and localized events like flood, black-

[1] http://www.tweepy.org

[2] a popular term of twitter messages

outs, and so forth. Besides that, it represents situations on moderate or high risk of people's lives and destruction of the infrastructure. In general, it can identify evacuations, damages, despair, calls for help, and so on. Such situations are not punctual and evolve very large number of affected.

To create the semantic models of our classes (i.e., the textual model that should be used as an input for the machine learning technique), we searched for news on renowned online newspapers (e.g., The New York Times and BBC News) to extract terms which are often used to detail information on disasters such as hurricanes. After analyzing 40 news, a set of 104 terms (e.g., blackout, flood, landfall) were extract and classified into the two classes based on the context in which the term appear, i.e., news to prepare the population for some event, or news announcing that something has happened. Both models were evaluated and refined with experts in crisis management

1.3.4. Machine Learning Technique Selection

Comparing related works in which tweets classification and sentiment analysis were performed, we considered a set of machine learning techniques which are frequently used in this context [Neppalli et al. 2017]. In this project, Naive Bayes was chosen because this technique does not consider the context of the context and every word in a message is independent. Such assumption is essential since we are using as a model a set of independent terms.

1.3.5. Implementation

The service was implemented considering all the results obtained in the previous steps. Such implementation was performed in Java language[3] through the Eclipse Mars[4] environment. The service is composed of four classes: (i) a class that connects to the twitter API, collects the data and sends the data to a SQL[5] database; (ii) a class responsible for communicating with the database; (iii) a class for the implementation of the machine learning technique; and (iv) a central class to manage the service.

1.4. Results and Discussions

The proposed service was developed to be a mediator between social networks and emergency responders, as shown in Figure 1.1. In this project, we implemented the internal structure of the service, and its connection to twitter interface.

Regarding the detection of the hurricane, we used the semantic models to classify the 124874 collected messages into two categories: (i) it is an occurrence of a disaster; (ii) it is not an occurrence of a disaster. Figure 1.2 presents the distribution of the classified messages in the period analyzed. The x-axis indicates the dates when the messages were posted, the y-axis indicates the number of messages that were classified as "is the occurrence of an event" on the respective day. As can be seen, results indicate a peak of messages in the period of September 8th to 10th, which suggests that our service can identify that some event was happening in such period of time. Considering these results,

[3]https://java.com/

[4]https://www.eclipse.org

[5]www.microsoft.com/SQL

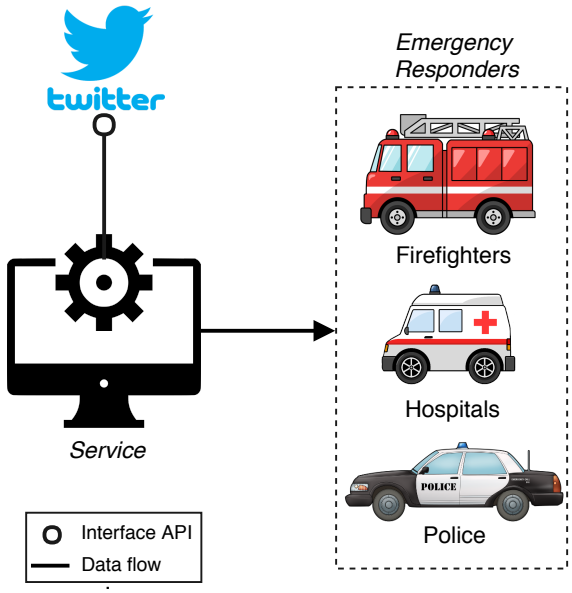

Figure 1.1. Data flow.

it is important to highlight that the hurricane Irma hit the analyzed region in the same period (i.e., from Sept 8th to 10th). Hence, it is possible to confirm the system made a correct classification of occurrence of disaster events. Besides that, we compared the results of such classification with a manual classification, to identify whether the system really learned from the semantic of the messages or if was just guessing the classification. As results, we obtained an accuracy of 76%, that indicates that the model is satisfactory at identifying the occurrence of disasters.

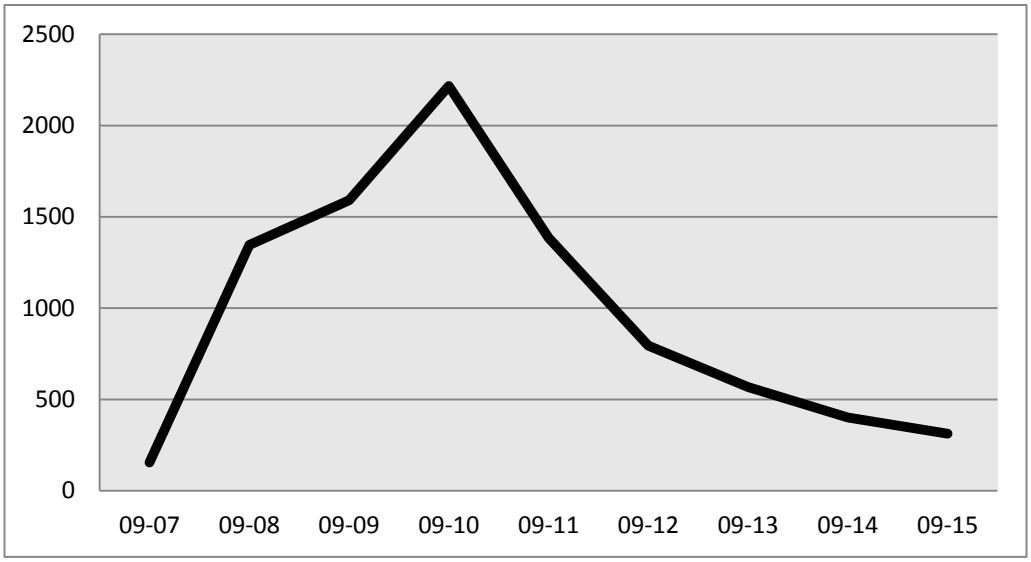

Figure 1.2. Amount of messages detected through the days.

To achieve the second goal, which aims at classifying messages according to their impact levels, we used the model to train the Naive Bayes technique in the full dataset.

Naive Bayes classified messages in low impact and high impact. Then, we performed a 10-fold cross validation [Kohavi et al. 1995] to test the precision and accuracy of such classification. Table 1.1 presents our results. The first and second columns present the classification precision for the low impact and high impact classes, respectively. The third column presents the average precision, while the last one presents the classification accuracy. As specified in the 10-cross validation definition, each row presents a classification alternating nine sub sets of data for training and one for testing [Kohavi et al. 1995].

	Observation Precision	**High Precision**	**Average Precision**	**Accuracy**
1	0.55	0.56	0.55	0.56
2	0.56	0.67	0.61	0.59
3	0.55	0.67	0.61	0.52
4	0.55	0.56	0.55	0.54
5	0.55	0.67	0.61	0.56
6	0.55	0.56	0.55	0.56
7	0.55	0.44	0.50	0.55
8	0.57	1,00	0.78	0.54
9	0.54	0.67	0.60	0.56
10	0.55	0.89	0.72	0.55
Total Average	0.55	0.67	0.61	0.55

Table 1.1. 10-fold cross validation

Results indicate that the precision of the low impact class was 55%, while the high impact class was 67%. However, the overall accuracy tends to the low impact class because it is the class with more observations, which is expected since high impact messages should be rare (i.e, an emergency event should not happen all the time). Thus, as an overall result, the service classified correctly 55% of the messages.

One of the major challenges faced in this development was the unbalanced classes. The proportion of low and high impact messages is from 5% to 95%. Such proportion generally implies in difficulties for generating a good model. However, we tried to mitigate this difficulty by creating a model based on terms extracted from online newspapers. Besides that, this proportion could cause the model to guess all messages into the class with the greatest proportion, in order to get most of the information correctly. Fortunately, due to the normalization applied in the Naive Bayes technique, this "guessing" did not happen in our results, since we still obtained results for the high class better than for the observation class.

Moreover, one problem related to choose twitter as a source of information is that the amount of text that can be typed by the user is small, causing them to use a variety of resources to send more complete information to their followers. One of the ways to send more information, is with the use of expressions and grammatical reductions of the language, for example the term *"for you"* can be reduced to *"4u"*. This type of analysis can hardly be performed by the classifier and so some accuracy points can be lost. The use of hyperlinks, images and different non textual elements can also impact the results.

However, the complexity to consider these features in the analysis increases exponentially.

1.5. Conclusions

In this project, we developed a service to assist emergency responders during a natural disaster. The developed service uses machine learning techniques and data provided by social network to detect that an event is occurring, and classify the messages into two different levels: high and low impact, so the responders could know where assistance is most needed.

As a result considering the Hurricane Irma, we obtained a precision of 76% for disaster detection, 55% for low impact messages classification, and 67% for high impact messages classification. Although not as high, such results are already satisfactory as a first approach in this perspective, and important for effective relief and even to save lives.

As future works, we intend to compare different machine learning techniques to verify whether they could improve the accuracy of our results. In addition, a deeper evaluation must be conducted as we are dealing with a critical domain. Finally, partnership with responsible entities should be considered to test the applicability of this approach in a real environment.

References

[Ashktorab et al. 2014] Ashktorab, Z., Brown, C., Nandi, M., and Culotta, A. (2014). Tweedr: Mining twitter to inform disaster response. In *ISCRAM*.

[Castillo et al. 2011] Castillo, C., Mendoza, M., and Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM.

[Chowdhury et al. 2013] Chowdhury, S. R., Imran, M., Asghar, M. R., Amer-Yahia, S., and Castillo, C. (2013). Tweet4act: Using incident-specific profiles for classifying crisis-related messages. In *ISCRAM*.

[Hui et al. 2012] Hui, C., Tyshchuk, Y., Wallace, W. A., Magdon-Ismail, M., and Goldberg, M. (2012). Information cascades in social media in response to a crisis: a preliminary model and a case study. In *Proceedings of the 21st International Conference on World Wide Web*, pages 653–656. ACM.

[Imran et al. 2013a] Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P. (2013a). Extracting information nuggets from disaster-related messages in social media. In *ISCRAM*.

[Imran et al. 2013b] Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P. (2013b). Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1021–1024. ACM.

[Kohavi et al. 1995] Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA.

[Kwak et al. 2010] Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM.

[Kyng et al. 2006] Kyng, M., Nielsen, E. T., and Kristensen, M. (2006). Challenges in designing interactive systems for emergency response. In *Proceedings of the 6th conference on Designing Interactive systems*, pages 301–310. ACM.

[Latonero and Shklovski 2010] Latonero, M. and Shklovski, I. (2010). "respectfully yours in safety and service emergency:" management & social media evangelism.

[Li et al. 2015] Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., Squicciarini, A. C., and Tapia, A. H. (2015). Twitter mining for disaster response: A domain adaptation approach. In *ISCRAM*.

[Mendoza et al. 2010] Mendoza, M., Poblete, B., and Castillo, C. (2010). Twitter under crisis: Can we trust what we rt? In *Proceedings of the first workshop on social media analytics*, pages 71–79. ACM.

[Neppalli et al. 2017] Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A., and Stehle, S. (2017). Sentiment analysis during hurricane sandy in emergency response. *International Journal of Disaster Risk Reduction*, 21:213–222.

[Palen et al. 2010] Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., and Grunwald, D. (2010). A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. In *Proceedings of the 2010 ACM-BCS visions of computer science conference*, page 8. British Computer Society.

[Phillips 2005] Phillips, D. (2005). Texas 9-1-1: Emergency telecommunications and the genesis of surveillance infrastructure. *Telecommunications Policy*, 29(11):843–856.

[Starbird et al. 2010] Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S. (2010). Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 241–250. ACM.

[Wex et al. 2014] Wex, F., Schryen, G., Feuerriegel, S., and Neumann, D. (2014). Emergency response in natural disaster management: Allocation and scheduling of rescue units. *European Journal of Operational Research*, 235(3):697–708.

[Yang et al. 2013] Yang, L., Yang, S.-H., and Plotnick, L. (2013). How the internet of things technology enhances emergency response operations. *Technological Forecasting and Social Change*, 80(9):1854–1867.

[Zareie and Sheikhahmadi 2017] Zareie, A. and Sheikhahmadi, A. (2017). A hierarchical approach for influential node ranking in complex social networks. *Expert Systems with Applications*.

11. Optimization for Hand Recognition

Chapter 11

Multi-Objective Optimization for Hand Posture Recognition

Sérgio F Chevtchenko and Valmir Macário

Abstract

Hand gestures are becoming increasingly popular in several application, such as smart houses, games, vehicle infotainment systems, kitchens and surgery rooms. An effective human-computer interaction system should aim for both good recognition accuracy and speed. A benchmark database is used for performance evaluation. It contains 2425 images with variations in scale, illumination and rotation. Several common image descriptors, such as Fourier, Zernike moments, pseudo-Zernike moments, Hu moments, complex moments and Gabor features are comprehensively compared by accuracy and speed for the first time. Experimental results showed good recognition rate, using a descriptor with low computational cost and reduced size. A real-time gesture recognition system based on the proposed descriptor is constructed and evaluated.

1.1. Introduction

Static hand gestures, also known in the literature as hand postures, can provide an intuitive human-computer interaction (HCI), in addition to being a non-verbal form of communication between humans. A hand gesture recognition system can be used for an effective and natural human-machine interaction [Wang and Wang, 2008]. It can also be beneficial for the deaf or hearing impaired, who mostly use gestures for communication.

The purpose of this work is to demonstrate a novel combination of features and optimization techniques for fast and accurate hand posture recognition. The proposed methodology has not been used before for hand posture recognition. A multi-objective genetic algorithm (NSGA-II) is designed to improve accuracy, minimize the number of features and simplify connections in an artificial neural network (ANN). Extensive experimentation is conducted with a hand posture database [Barczak et al., 2011] and the proposed method is compared with existing ones in terms of recognition accuracy, computational cost and feature vector size. Finally, a real-time gesture recognition system based on the proposed descriptor is implemented and evaluated.

The remainder of this Chapter is organized as follows. Background information and related work are provided in Section 1.2. The details of the proposed method for hand detection are described in Section 1.3. Experimental results and comparisons are demonstrated in Section 1.4 and conclusions are presented in Section 1.5.

1.2. Background

This section provides basic information about hand posture recognition and multi-objective optimization. Related work on hand posture recognition is also reviewed.

1.2.1. Hand Posture Recognition

A hand posture recognition system consists essentially of the following steps: (a) image acquisition; (b) hand segmentation; (c) feature extraction; (d) hand posture classification. The image descriptors, after extraction from a segmented frame, are used to construct the feature vector. In the proposed method, an evolutionary algorithm is used to iteratively select the best combination of features so that the following two goals are met: optimal recognition accuracy and minimum computational cost. The resulting feature vector is less complex and therefore is less likely to cause overfitting.

Similarly to the feature extraction step, gesture classification has to be both fast and accurate. A multilayer perceptron (MLP) network is used to recognize the hand posture, and the feature vector is input to this network. However, the network's performance within a given training period is dependent on its structure. A small network may have limited processing capability, while a large one may be too slow and may have redundant connections. Thus, the same evolutionary algorithm is also used to select the optimal layout for the neural network.

1.2.2. Multi-Objective Optimization

Multi-objective problems occur when an optimal decision is a trade-off between two or more conflicting objectives. Feature selection is inherently a multi-objective problem, where any given solution can be evaluated in terms of the complexity of selected features and the corresponding accuracy. The objectives are to minimize the complexity of the feature vector and to maximize accuracy.

In the feature selection problem addressed in this paper, one solution A is said to dominate another solution B if the following criteria are met:

- A has lower feature complexity than B and/or yields better accuracy.
- A is not worse than B, either in terms of complexity or of accuracy.

1.2.3. Related Works

Numerous methods have been recently developed in order to classify hand gestures successfully. [Hasan and Abdul-Kareem, 2014] and [Badi et al., 2014] present a hand gesture recognition system. A total of 6 different gestures is classified with an 86.38% recognition rate, using a multilayer perceptron and complex moments. In this paper, complex moments are compared against other common image descriptors with respect to accuracy and speed.

[Aowal et al., 2014] extensively test discriminative Zernike moments (DZMs) and compare them with standard Zernike moments, principal component analysis (PCA) and Fourier descriptors (FDs). [Ng and Ranganath, 2002] develop a real-time system that uses Fourier descriptors to represent hand blobs. An RBF network is used to classify hand poses. Hu moments have been compared with Zernike moments for general object recognition [Sabhara et al., 2013] as well as for static hand gestures [Otiniano-Rodríguez et al., 2012]. In general, Zernike moments are found to be more accurate than Hu moments [Otiniano-Rodríguez et al., 2012] [Sabhara et al., 2013]. [Guo et al., 2015] propose an improved version of Zernike moments, measuring accuracy as well as computational cost on a common static hand gesture database, which this paper also uses. [Avraam, 2014] proposes a combination of graph-based characteristics and appearance-based descriptors such as detected edges for modeling static gestures. This method is tested on 10 numerical gestures from the Massey database [Barczak et al., 2011]. An average recognition rate of 83% with a standard deviation of 10% is achieved. [Cambuim et al., 2016] propose and efficient hand posture recognizer implemented on an FPGA. Gestures are converted into binary images and classified by an Extreme Learning Machine (ELM) neural network. On an FPGA, the system achieves 97% recognition rate at 6.5 ms per image. A sequential forward feature selection technique (SFS) is implemented by [Rasines et al., 2014] to select the most promising feature set among a large variety of features, such as Gabor features, histogram of oriented gradient (HOG), Shannon entropy, Hu moments and others. The results are compared to [Avraam, 2014], thereby improving the average recognition rate on 10 gestures of the same database to 97.5%. More specifically, gesture "7" is recognized with an accuracy of 75%, while other gestures are recognized without error. In the present study, the results are presented for the entire set of gestures from the Massey dataset, including 26 letters and 10 digits.

Given the importance of choosing the best feature set to represent hand postures and the computational cost of the best classifier to perform this task in real time, this paper compares several commonly used features, namely Zernike and pseudo-Zernike moments, Fourier descriptors, complex moments, Hu features and Gabor features. These image descriptors are compared with each other based on recognition accuracy and computational cost. Given that feature selection is a multi-objective problem, this paper uses a well-known multi-objective evolutionary algorithm (NSGA-II) in order to select the best combination of features and to optimize the performance of the neural network.

Although artificial neural networks and genetic algorithms have been used separately in this problem, this study is unique as it combines both. Also, as far as the authors know, the feature descriptors tested in this paper are compared for the first time with respect to accuracy and recognition speed, thereby providing a good starting point for future studies.

1.3. Methodology

The following feature descriptors are implemented and evaluated in this work: Hu moments, complex moments, Zernike moments, pseudo-Zernike moments, Fourier descriptors and Gabor features. Combinations of these features are also considered.

A configurable three-layer perceptron network is used to classify the feature vec-

tor. The network has l inputs, n neurons in the hidden layer and m outputs. A multi-objective evolutionary algorithm known as NSGA-II (Nondominated Sorting Genetic Algorithm) is adapted for this application. NSGA-II is a popular algorithm introduced by [Deb et al., 2002]. It has recently been used for a variety of multi-objective optimization problems but it is proposed for the first time for hand posture recognition in this paper.

1.4. Experimental Evaluation

There are several strategies for experimental design which involve multi-objective feature selection and presentation of the results. In this paper, we follow the strategy found in [Xue et al., 2013], [Hamdani et al., 2007] and [Ishibuchi et al., 2006]. In order to increase statistical robustness, the results are averaged across a total of 50 experiments. The dataset is always divided randomly into training (80%) and testing (20%) sets.

The Massey University Gesture Dataset [Barczak et al., 2011] is a well-known American Sign Language (ASL) database. It was published in 2011 and contains 2425 color images of 36 classes (26 letters and 10 numeric gestures). With the exception of [Guo et al., 2015] and [Kumar et al., 2014], other related studies provide results for only a part of this dataset [Aowal et al., 2014, Avraam, 2014, Rasines et al., 2014]. Although our experiments were carried also with parts of the Massey dataset, the results are presented for the entire database.

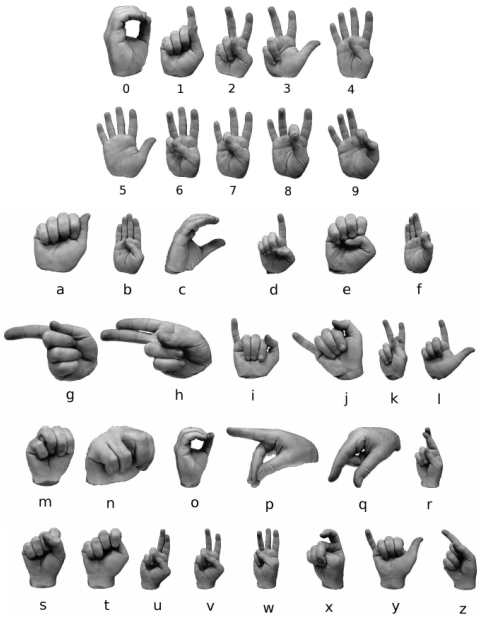

Figure 1.1: The Massey hand posture database [Barczak et al., 2011].

Figure 1.1 shows samples of different classes. Note that some gestures are rather difficult to distinguish. For example “a” and “e”, “d” and “1”, “m” and “n” or “i” and “j”. Furthermore, the database contains variations in scale, illumination and rotation. In order to minimize the influence of scale and increase computation speed, the images are normalized to 64×64 pixels prior to the feature extraction step.

1.4.1. Implementation

The feature extraction methods defined in the previous section are implemented as stated below. The combinations of different feature descriptors are presented here for the first time in the context of hand posture recognition.

- **HU** – Implemented as in [Otiniano-Rodríguez et al., 2012]. An unmodified OpenCV implementation of Hu moments is used, resulting in a vector of 7 features.

- **FD** – Implemented as in [Ng and Ranganath, 2002] and [Zhang and Lu, 2001]. Fourier descriptor using centroid distance function, extracted from a contour with 120 points.

- **FD2** – Implemented as in [Ng and Ranganath, 2002] and [Zhang and Lu, 2001]. The absolute value of a Fourier descriptor using complex coordinates is extracted from a contour with 120 points.

- **CM** – 10 complex moments, implemented as in [Hasan and Abdul-Kareem, 2014] and [Badi et al., 2014].

- **PZM** – Implemented as in [Chong et al., 2003]. Absolute value of complex pseudo-Zernike moments up to order 10, ignoring the first two moments, resulting in a vector of 64 features.

- **ZM** – Implemented as in [Otiniano-Rodríguez et al., 2012]. Zernike moments up to order 10, ignoring the first two moments, resulting in a vector of 34 features.

- **ZM_25** – Zernike moments up to order 25, ignoring the first two moments. Implemented as in [Guo et al., 2015].

- $\mathbf{GB}_{\sigma=0.02}$ and $\mathbf{GB}_{\sigma=0.2}$ – 64 features based on a Gabor filter with the following parameters: $w = 64$, $\theta = 0$, $\lambda = 1$, $\gamma = 0.02$, $\psi = 0$. Two different values of standard deviation σ are used: 0.02 and 0.2.

- **GB2** – Concatenation of $\mathrm{GB}_{\sigma=0.02}$ (64 features) and $\mathrm{GB}_{\sigma=0.2}$ (64 features), resulting in a vector with 128 features.

- **GB2_HU** – Concatenation of GB2 (128 features) and HU (7 features), resulting in a vector with 135 features.

- **GB2_ZM** – Concatenation of GB2 (128 features) and ZM (34 features), resulting in a vector with 162 features.

- **GB2_HU_ZM** – Concatenation of GB2 (128 features), HU (7 features) and ZM (34 features), resulting in a vector with 169 features.

Table 1.1: Comparison with existing methods

Feature descriptor	Work	# of features	Accuracy (%)	CPU time (ms)	Strengths	Weaknesses
HU	Otiniano-Rodríguez et al. (2012)	7	37.02	0.70	Very fast and easy to implement	Low accuracy
CM	Badi et al. (2014), Hasan and Abdul-Kareem (2014)	10	48.09	2.10	Fast	Low accuracy
ZM	Otiniano-Rodríguez et al. (2012)	34	91.08	36.30	Good accuracy	High CPU time
ZM_25	Guo et al. (2015)	180	92.08	252.00	Good accuracy	Very high CPU time
GB2_HU	This paper	135	97.09	2.82	Low CPU time; high accuracy	More adjustable parameters
GB2_ZM	This paper	162	97.63	39.01	High accuracy	More adjustable parameters; high CPU time

1.4.2. Experimental Results

A benchmark database [Barczak et al., 2011] is used to evaluate several feature descriptors with respect to accuracy and computational cost. A detailed comparison is presented in Table 1.1, including main strengths and weaknesses found by means of experimentation. The feature vectors proposed in this paper outperform recent static gesture recognition techniques based on Zernike and complex moments. Gabor features are used to extract texture information from the image and are shown to be a good combination with Zernike moments for hand posture recognition. The main weakness of Gabor features is that the Gabor filter has parameters that have to be adjusted for the specific application. While combining Zernike and Gabor features results in a good recognition rate, the multi-objective analysis suggest that a combination of Gabor and Hu descriptors can achieve better performance in terms of both speed and accuracy.

1.4.3. Real-Time Gesture Recognition

In order to apply the results obtained by multi-objective analysis to a real-time gesture recognition system, we choose a single nondominated solution. The real-time system is implemented in C++, using the OpenCV library. It runs on a notebook with an Intel i5 processor. As previously stated, hand segmentation is a difficult problem and is usually treated separately. In order to simplify this step, the hand is filmed against a black background, and a black wristband is worn.

On average, once the frame is acquired, preprocessing (steps 5 to 8) takes 28ms, feature extraction (step 9) takes 2ms and the neural network (step 10) takes less than 1ms to recognize the feature vector. This means that, when using a common 30fps camera, a single frame can be processed and recognized before the next frame is acquired, resulting in no noticeable delay for the user. Note that no special effort is made to optimize the image processing steps, as this is not in the scope of this paper. Additionally, since the recognition step takes less than 1ms, a more advanced classifier, such as an ensemble of classifiers, can be used.

1.5. Conclusion

Recognition accuracy of up to 98.61% is achieved for 36 gestures, while using a feature vector with a low computational cost (extraction time of 2.82 ms for a 64×64 image). This feature vector is used to make a real-time gesture recognition system. Additional experimental results can be found in [Chevtchenko et al., 2018].

This work makes the following main contributions:

- several commonly used feature extraction methods are tested on a benchmark dataset and compared in terms of their accuracy and computational cost, for the first time.
- a multi-objective feature selection and classifier optimization is applied to hand posture recognition.
- a real-time gesture recognition system is implemented based on the optimized configuration returned by an adapted version of NSGA-II.

References

[Aowal et al., 2014] Aowal, M. A., Zaman, A. S., Mahbubur Rahman, S., and Hatzinakos, D. (2014). Static hand gesture recognition using discriminative 2D zernike moments. In *TENCON 2014-2014 IEEE Region 10 Conference*, pages 1–5. IEEE.

[Avraam, 2014] Avraam, M. (2014). Static gesture recognition combining graph and appearance features. *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, 3(2).

[Badi et al., 2014] Badi, H., Hussein, S. H., and Kareem, S. A. (2014). Feature extraction and ml techniques for static gesture recognition. *Neural Computing and Applications*, 25(3-4):733–741.

[Barczak et al., 2011] Barczak, A., Reyes, N., Abastillas, M., Piccio, A., and Susnjak, T. (2011). A new 2D static hand gesture colour image dataset for asl gestures.

[Cambuim et al., 2016] Cambuim, L. F., Macieira, R. M., Neto, F. M., Barros, E., Ludermir, T. B., and Zanchettin, C. (2016). An efficient static gesture recognizer embedded system based on elm pattern recognition algorithm. *Journal of Systems Architecture*, 68:1–16.

[Chevtchenko et al., 2018] Chevtchenko, S. F., Vale, R. F., and Macario, V. (2018). Multi-objective optimization for hand posture recognition. *Expert Systems with Applications*, 92:170–181.

[Chong et al., 2003] Chong, C.-W., Raveendran, P., and Mukundan, R. (2003). An efficient algorithm for fast computation of pseudo-Zernike moments. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(06):1011–1023.

[Deb et al., 2002] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.

[Guo et al., 2015] Guo, Y., Liu, C., and Gong, S. (2015). Improved algorithm for Zernike moments. In *Control, Automation and Information Sciences (ICCAIS), 2015 International Conference on*, pages 307–312. IEEE.

[Hamdani et al., 2007] Hamdani, T. M., Won, J.-M., Alimi, A. M., and Karray, F. (2007). Multi-objective feature selection with NSGA-II. In *International Conference on Adaptive and Natural Computing Algorithms*, pages 240–247. Springer.

[Hasan and Abdul-Kareem, 2014] Hasan, H. and Abdul-Kareem, S. (2014). Static hand gesture recognition using neural networks. *Artificial Intelligence Review*, 41(2):147–181.

[Ishibuchi et al., 2006] Ishibuchi, H., Nojima, Y., and Doi, T. (2006). Comparison between single-objective and multi-objective genetic algorithms: Performance comparison and performance measures. In *2006 IEEE International Conference on Evolutionary Computation*, pages 1143–1150. IEEE.

[Kumar et al., 2014] Kumar, V., Nandi, G. C., and Kala, R. (2014). Static hand gesture recognition using stacked denoising sparse autoencoders. In *Contemporary Computing (IC3), 2014 Seventh International Conference on*, pages 99–104. IEEE.

[Ng and Ranganath, 2002] Ng, C. W. and Ranganath, S. (2002). Real-time gesture recognition system and application. *Image and Vision computing*, 20(13):993–1007.

[Otiniano-Rodríguez et al., 2012] Otiniano-Rodríguez, K., Cámara-Chávez, G., and Menotti, D. (2012). Hu and Zernike moments for sign language recognition. In *Proceedings of international conference on image processing, computer vision, and pattern recognition*, pages 1–5.

[Rasines et al., 2014] Rasines, I., Remazeilles, A., and Iriondo Bengoa, P. M. (2014). Feature selection for hand pose recognition in human-robot object exchange scenario. In *Emerging Technology and Factory Automation (ETFA), 2014 IEEE*, pages 1–8. IEEE.

[Sabhara et al., 2013] Sabhara, R. K., Lee, C.-P., and Lim, K.-M. (2013). Comparative study of Hu moments and Zernike moments in object recognition. *SmartCR*, 3(3):166–173.

[Wang and Wang, 2008] Wang, C.-C. and Wang, K.-C. (2008). Hand posture recognition using Adaboost with SIFT for human robot interaction. In *Recent progress in robotics: viable robotic service to human*, pages 317–329. Springer.

[Xue et al., 2013] Xue, B., Zhang, M., and Browne, W. N. (2013). Particle swarm optimization for feature selection in classification: a multi-objective approach. *IEEE transactions on cybernetics*, 43(6):1656–1671.

[Zhang and Lu, 2001] Zhang, D. and Lu, G. (2001). A comparative study on shape retrieval using Fourier descriptors with different shape signatures. In *Proc. International Conference on Intelligent Multimedia and Distance Education (ICIMADE01)*.