

Classificação multi-rótulo de aspectos em avaliações online

Lucas Vinicius M. da Frota^{1,2}, Carlos Maurício S. Figueiredo^{1,2}

¹Núcleo de Computação – Universidade do Estado do Amazonas (UEA)
Manaus – AM – Brasil

²Samsung Ocean Center (OCEAN)
Manaus – AM – Brasil

{lvmdf.snf, cfigueiredo}@uea.edu.br

Abstract. *With the popularization of the Internet in recent years it has been possible to accumulate large amounts of unstructured data in the form of comments or reviews thus interpreting them can generate important information for any company that wants to improve its services or products. Therefore this article aims to create a machine learning model capable of inferring which topics are being mentioned within a sample of online review free text. Being shown in restaurant evaluation case study.*

Resumo. *Com o a popularização da internet nos últimos anos houve o acúmulo de grandes volumes de dados não estruturados na forma de comentários ou avaliações, de tal maneira que interpretá-los pode gerar informações fundamentais para qualquer empresa que deseja melhorar seus serviços ou produtos. Tendo isso em vista este artigo visa criar um modelo de aprendizagem de máquina capaz de inferir quais tópicos estão sendo mencionados dentro de uma amostra de texto livre de avaliação online. Isto é mostrado em estudo de caso de avaliação de restaurante.*

1. Introdução

Com o surgimento e a popularização da internet houve o acúmulo de grandes quantidades de dados, em sua maioria dados não estruturados como vídeo, áudio, imagens e textos por exemplo. Grandes serviços da atualidade utilizam dados desse tipo para entender melhor seus clientes, empresas como *Uber*, *AirBnB* e *Amazon* por exemplo, utilizam textos como uma forma de pedir *feedback* de seus produtos e serviços, gerando assim grandes massas de dados que se tornam muito difíceis de analisar manualmente. Portanto entender os assuntos que aos quais esses textos se referem se torna uma tarefa importante para o crescimento e melhoria de serviços como estes.

Recentemente muitos trabalhos têm explorado a área de detecção de aspectos no texto [Hu and Liu 2004, Poria et al. 2016, Souza et al. 2018], outros trabalhos tentam além de extrair os aspectos, inferir suas polaridades [Pontiki et al. 2016, Wang et al. 2016, Kiritchenko et al. 2014]. Assim além de entender sobre quais aspectos uma avaliação está se referindo também é possível ver se o comentário em questão é positivo ou negativo.

No entanto, é preciso ter em mente que muitos textos possuem mais de um aspecto e portanto mais de uma polaridade associada, assim em uma frase como "*A comida é ótima, mas o ambiente é terrível*" é possível extrair os aspectos comida e ambiente, e

suas respectivas polaridades, positiva e negativa. O problema da análise de sentimentos baseada em aspectos pode desta maneira ser dividido em duas tarefas: (i) Detecção dos aspectos presentes no texto; (ii) Classificação da polaridade de cada aspecto.

Neste trabalho buscamos solucionar a primeira tarefa, criando assim um modelo de aprendizado de máquina capaz de inferir quais dentre cinco aspectos estão sendo mencionados dentro de um texto avaliativo no contexto de um restaurante.

O restante deste trabalho está organizado da seguinte maneira: A seção 2 mostra trabalhos que relacionados a análise de sentimentos baseada em aspectos; A seção 3 explica a abordagem que será utilizada neste trabalho; Na seção 4 são mostrados os experimentos e os resultados obtidos; A seção 5 mostra as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

A análise de sentimentos baseada em aspectos é uma subárea da mineração de textos que visa analisar grandes volumes de dados textuais, como avaliações em lojas virtuais ou textos em redes sociais.

Trabalhos anteriores tentaram resolver a tarefa (i) através de uma abordagem de classificação multi-classe, [Souza et al. 2018, Pontiki et al. 2016] de tal maneira que dado um texto este só poderia ser classificada como possuindo um único aspecto, esta abordagem acarreta perdas de informação na análise dos textos, pois muitas vezes o mesmo texto se refere a mais de um assunto ao mesmo tempo. Assim, ao utilizar esta estratégia alguns assuntos que estão de fato presentes nos dados podem ser ignorados gerando assim análises que não mostram a realidade dos dados.

Outra abordagem explorada foi a criação de modelos específicos para cada aspecto como o trabalho de [Kiritchenko et al. 2014]. Neste trabalho foram gerados cinco modelos através do algoritmo *Support Vector Machine*(SVM) onde cada um realiza uma classificação binária a fim de inferir a presença ou ausência de um determinado tópico do texto, assim é possível gerar um classificação multi-categoria, onde cada amostra pode pertencer a zero ou mais classes. Porém esta abordagem se mostra mais complexa uma vez que aqui é necessário gerar uma quantidade de modelos igual a quantidade dos aspectos que se deseja prever.

De maneira diferente dos artigos citados acima, este trabalho utiliza uma abordagem multi-rótulo para determinar quais aspectos estão presentes em um texto no domínio de avaliações de restaurantes através de um único modelo de aprendizado de máquina.

3. Abordagem Proposta

A abordagem aqui proposta consiste na utilização de *word embeddings* para representar as palavras contidas no texto, com o intuito de tornar a tarefa de classificar quais aspectos estão presentes nele mais simples. Estas, por sua vez, será realizada por uma rede neural, nesta seção estes dois elementos serão explicados com mais detalhes, juntamente com as ferramentas utilizadas.

3.1. *Word embedding*

Word embedding é uma abordagem que tem sido amplamente empregada em trabalhos recentes relacionados a classificação de texto [Poria et al. 2016, Kim 2014]. Por isso ela

foi utilizada no presente trabalho. Nesta abordagem cada palavra é representada por um vetor de 20 posições. No modelo proposto, a primeira camada utilizada foi de *embedding*.

No conjunto de dados, o texto com o maior número de palavras possui aproximadamente 70, por isso a entrada da rede neural foi um vetor de mesmo tamanho, onde a ordem original das palavras foi mantida. Nos casos onde a entrada possuía menos de 70 palavras os demais espaços foram preenchidos com um identificador de ausência.

3.2. Classificador

Após a entrada passar pela primeira camada os identificadores numéricos equivalentes as palavras são convertidos em vetores densos que são mais significativos para as redes neurais, formando assim um vetor de *word embeddings*, ou seja uma matriz de dimensão 70x20. Em seguida essa matriz é convertida em um vetor plano que assim é passado como entrada para a rede neural que realiza a classificação.

O classificador por sua vez é composto por 3 camadas totalmente conectadas, sendo elas camadas com 64, 32 e 5 neurônios que utilizam na primeira a função de ativação *softmax* e nas demais a função de ativação *sigmoid*.

3.3. Ferramentas

Para a construção deste modelo foi utilizada a linguagem de programação *Python*¹ em sua versão 3.6.7, juntamente com o *framework* de abstração de *deep learning* Keras² (versão 2.2.4), utilizando como *backend* o TensorFlow³(versão 1.13.1).

4. Experimentos e Resultados

Esta seção mostra a base de dados utilizada junto com uma breve análise de algumas de suas características, os experimentos realizados, e finalmente os resultados obtidos.

4.1. Base de dados

Neste trabalho extraímos os aspectos relativos a avaliações de restaurantes. Para isso, utilizamos a mesma base de dados utilizada no trabalho de [Poria et al. 2016]. Esta pode ser encontrada no site SemEval⁴. Esta base possui 3044 amostras, nela foi realizada uma partição do tipo *hold out* na qual 70% dos dados foram utilizados para treinar o modelo e 30% foram utilizados para a realização dos testes. Desta maneira, foi possível avaliar o modelo com amostras nunca antes vistas por ele. É importante ressaltar que esta base contém textos apenas em inglês.

Nessa base, cada amostra pode conter entre um e cinco aspectos, sendo eles: serviço, comida, anedotas, preço e ambiente. A recorrência de cada um desses aspectos no conjunto de dados está ilustrada na Figura 1.

Pelo gráfico é possível notar que as duas classes mais presentes nos textos são comida e anedotas. Isso, provavelmente, se deve ao fato de que a comida e as experiências

¹<https://www.python.org/>

²<https://keras.io/>

³<https://www.tensorflow.org/>

⁴<http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools>

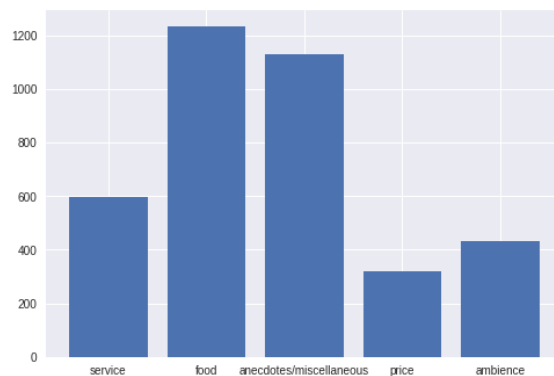


Figura 1. Distribuição dos aspectos na base de dados

vivenciadas no restaurante são pontos centrais para o cliente deste tipo de estabelecimento, também é possível ver que comentários a respeito do preço são os menos comuns dentro do conjunto.

Outra informação relevante sobre o conjunto de dados é a quantidade de aspectos presentes em cada amostra. A grande maioria das amostras (81,1%) se refere a apenas um aspecto, enquanto 16% se refere a 2 aspectos, 2,7% se refere a 3, 0,2% se refere a 4 e nenhuma das amostras cita 5 aspectos distintos. Essas proporções se mantêm muito similares nas partições de treino e teste.

4.2. Treinamento

No treinamento deste modelo foi utilizado o otimizador Adam e a função de perda *binary crossentropy*. É importante salientar que esses parâmetros foram escolhidos de maneira empírica. O modelo foi treinado ao longo de 100 épocas.

4.3. Métricas de avaliação

Por se tratar de um problema de classificação onde cada amostra pode pertencer a mais de uma classe ao mesmo tempo é necessário ter em mente que há duas maneiras principais de medir o desempenho do modelo que tenta resolver esse problema: (i) Medir os acertos exatos, ou seja só é contabilizado um acerto quando o modelo infere de maneira correta a presença ou a ausência de todas as cinco classes as quais a amostra pode pertencer, de tal modo que caso um desses seja inferido de maneira incorreta toda a previsão sera contabilizada como incorreta; (ii) Medir os acertos parciais, ou seja contar a quantidade de acertos de uma previsão mesmo que esta não seja completamente igual à saída desejada, desta forma caso algum aspecto específico seja inferido de maneira incorreta os demais não serão desconsiderados. Também é importante lembrar que neste problema acertar tanto a ausência quanto a presença de um aspecto é igualmente importante.

Dado que a estrutura da saída da rede é um vetor de números entre 0 e 1, onde cada posição representa a probabilidade de um determinado aspecto estar presente naquele texto, foi necessário escolher um limiar para definir a presença ou ausência do aspecto. Aqui ele foi definido como 0,5, pois esta é a média entre 0 e 1. Assim dada uma determinada saída do modelo cada um dos valores contidos nela será interpretado como "possui o aspecto" se maior que 0,5 e caso contrário será interpretado como "não possui o aspecto".

4.4. Resultados

Nas subseções seguintes utilizaremos o conjunto de testes descrito no início desta seção para a realização os testes.

4.4.1. Acertos exatos

Considerando apenas os acertos exatos, foi possível alcançar 446 acertos dentro de um universo de 914 amostras, totalizando assim 48% de acerto quando medido desta forma. É interessante notar que 435 acertos foram relativos a textos que contêm apenas um aspecto, os demais 11 acertos são relativos a textos com dois aspectos presentes. Sendo assim, não houve nenhum acerto exato ao tentar prever textos com mais de dois aspectos. Essa baixa taxa de acerto em textos com mais de dois aspectos provavelmente se deve ao fato de existirem poucas amostras no conjunto de dados com estas características, menos de 3%. Por isso, é possível imaginar que com um conjunto de dados mais balanceado seria possível alcançar uma taxa de acerto exato mais alta para este problema.

4.4.2. Acertos parciais

Neste teste medimos a capacidade do modelo de classificar individualmente a presença ou ausência de cada aspecto dentro de um texto. Desta forma cada texto do conjunto de teste foi interpretado não mais como uma única amostra, mas sim como cinco amostras. Desta maneira se a previsão estivesse totalmente correta esta seria contabilizada como cinco acertos e não mais como apenas um, como feito no teste anterior. Ao testar desta maneira foi possível obter uma taxa de 83% de acerto. Quando medidas de maneira individual as acurácias de cada aspecto geraram os resultados mostrados na Figura 2.

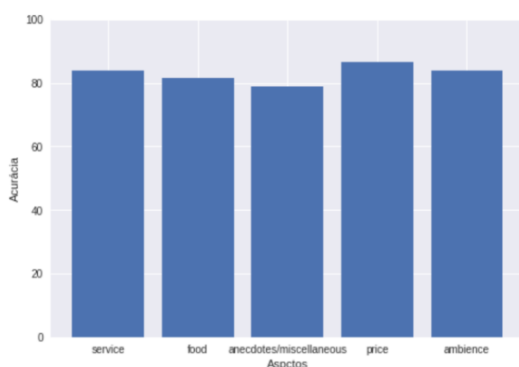


Figura 2. Acurácia de cada aspecto

Tabela 1. Métricas acurácia, precisão e revocação

	Acurácia	precisão	Revocação	F1
Serviço	86%	56%	64%	60%
Comida	83%	77%	81%	79%
Anedotas	79%	77%	70%	73%
Preço	82%	16%	17%	17%
Ambiente	83%	3%	18%	5%

Como pode ser observado na Figura 3, foi possível obter uma acurácia próxima de 80% para todos os aspectos. Entretanto, ao analisar outras métricas foi possível ver que apesar da alta acurácia, este modelo ainda apresenta resultados não muito satisfatórios. A tabela 1 mostra além da acurácia a precisão, revocação e o *F1 score* de cada um dos aspectos. Através da análise dessas métricas foi possível notar que os aspectos que possuem pontuações menores foram os mesmos que possuíam menos dados, enquanto que os aspectos mais comuns no conjunto de dados possuem pontuações maiores, como a comida

por exemplo onde o F1 *score* foi de 79%. Assim é possível imaginar que com um *dataset* mais balanceado seria possível alcançar valores mais compatíveis com a acurácia de total.

5. Conclusões e diretrizes futuras

Neste trabalho, foi possível atingir o objetivo inicial de gerar um modelo capaz de realizar uma classificação multi-rótulo identificando assim, de maneira simultânea quais dos cinco aspectos relativos a avaliações de restaurantes estão presentes ou não em um determinado texto.

Apesar de ter sido capaz de prever perfeitamente apenas 48% do conjunto de teste esse modelo se provou capaz de generalizar bem os aspectos que se encontram bem representados dentro do conjunto de treino, chegando a classificar 83% dos aspectos da base de teste corretamente.

Além disso é importante ter em mente que o presente estudo se limita a realizar a tarefa proposta utilizando textos escritos na língua inglesa no contexto de restaurantes, é necessário portanto realizar estudos futuros analisando a aplicação do modelo proposto em outros idiomas e contextos.

Referências

- Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- Kiritchenko, S., Zhu, X., Cherry, C., and Mohammad, S. (2014). Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A.-S., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 19–30.
- Poria, S., Cambria, E., and Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49.
- Souza, B. A., Menezes, A. A., Figueiredo, C. M., Nakamura, F. G., and Nakamura, E. F. (2018). Detecção de categorias de aspectos utilizando redes neurais profundas em avaliações online. In *7º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)*, volume 7. SBC.
- Wang, Y., Huang, M., Zhao, L., et al. (2016). Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615.