

Desenvolvimento de um Modelo Semântico para Recomendação Baseado em Grafos

Hugo Guércio Fernandes¹, Victor Ströele¹

¹Departamento de Ciência da Computação
Universidade Federal de Juiz de Fora
Juiz de Fora – MG – Brazil

{hugo.guercio,victor.stroele}@ice.ufjf.br

Abstract. *Recommender Systems aims to help a user or groups of users in identifying the most relevant items based on their needs. The items may have different characteristics and can be services, products or miscellaneous information. With the growth of data amount, recommender systems are being increasingly studied because it is increasingly difficult to find the desired information due to the available alternatives. This paper proposes and develop a model for a recommender system based on bipartite graphs, formed by semantic information of users and items. This recommendation model was evaluated based on two Proof of Concepts (PoC).*

Resumo. *Os sistemas de recomendação tem como objetivo auxiliar um usuário, ou grupos de usuários, na identificação de diferentes itens relevantes às suas necessidades. Com a crescente quantidade de dados, os sistemas de recomendação estão sendo muito estudados, pois é cada vez mais difícil encontrar a informação desejada devido a grande quantidade de alternativas disponíveis. Neste trabalho é proposto e desenvolvido um modelo para um sistema de recomendações baseado em grafos bipartidos, composto por informações semânticas de usuários e itens. Este modelo de recomendação foi avaliado com base em duas provas de conceitos (PoC).*

1. Introdução

O desenvolvimento e uso de tecnologias, que dão apoio à interação social, cresce rapidamente. Com isso novas formas de comunicação, trabalho e entretenimento são usadas por bilhões de pessoas gerando redes complexas que representam essas interações entre elas.

Com o desenvolvimento das tecnologias e aumento do volume de dados é necessário auxiliar os usuários a encontrarem o que, de fato, é do interesse deles. Assim, os sistemas de recomendação tentam identificar os recursos que sejam aderentes aos interesses dos seus usuários em diferentes cenários, como: análise de mercado, recomendação de recursos educacionais e de apoio à pesquisa, recomendação de produtos, controle de produção ou retenção de consumidores [Han and Kamber 2006].

As recomendações, na maioria das vezes, são feitas a partir do conhecimento extraído de dados históricos dos usuários e itens. Casos bem conhecidos de recomendações podem ser exemplificados pela sugestão de itens na Amazon.com como analisado em [Oestreicher-Singer and Sundararajan 2010], e em filmes na Netflix.com.

O objetivo deste trabalho é propor e desenvolver um modelo de recomendação de itens a pesquisadores baseado em grafos bipartidos. Esses grafos são compostos por pesquisadores, vídeos e informações semânticas extraídas através da análise dos pesquisadores e dos vídeos.

Como resultado do modelo proposto foram geradas recomendações de vídeos para alguns pesquisadores com o intuito de avaliar a proposta. Para isso, foram extraídas informações de uma rede social científica [Ströele et al. 2013], composta pelos relacionamentos de coautoria entre os pesquisadores presentes na DBLP (Bibliografia de Ciência da Computação).

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta a metodologia adotada neste trabalho; a Seção 3 tem como objetivo fundamentar os conceitos utilizados; a Seção 4 apresenta alguns trabalhos relacionados; na Seção 5 é apresentado o modelo proposto para o problema; já na Seção 6 são apresentados os resultados encontrados seguidos pela Seção 7 que apresenta uma avaliação do modelo proposto. Por fim, são apresentadas as considerações finais na Seção 8.

2. Metodologia

A metodologia que este trabalho utilizou possui: (1) pesquisa bibliográfica para identificar problemas e abordagens da área de recomendação, identificação dos tipos de recursos a serem recomendados e identificação de trabalhos relacionados; (2) proposta e desenvolvimento de um modelo para recomendação através de informações semânticas; e (3) avaliação do modelo desenvolvido através de duas provas de conceitos.

A partir da pesquisa bibliográfica foi possível identificar os aspectos mais relevantes das soluções relacionadas e utilizar o resultado dessa pesquisa para fundamentar a construção do modelo. Para validação do resultado foi utilizada uma rede de coautoria extraída do repositório DBLP¹. Assim, os usuários são pesquisadores de uma rede social científica e os itens recomendados foram vídeos extraídos do Youtube. Para avaliação do modelo desenvolvido foram realizadas duas provas de conceitos com alguns pesquisadores dessa rede social.

3. Recomendação de conteúdo com base em redes sociais

Os Sistemas de Recomendação representam uma área de pesquisa importante devido a abundância de aplicações práticas, que ajudam os usuários a lidar com a grande quantidade de dados presentes em diversas plataformas. De acordo com [Adomavicius and Tuzhilin 2005] os sistemas de recomendação vem sendo, em geral, separados em três categorias: baseada em conteúdo, colaborativa e híbrida. Neste trabalho foi utilizada uma abordagem baseada em conteúdo.

A abordagem de filtragem baseada em conteúdo tem suas raízes nos processos de recuperação da informação [Baeza-Yates et al. 1999]. Neste tipo de filtragem é realizada uma descrição dos usuários e itens com base em suas características. Após essa caracterização é realizada uma comparação entre as descrições, verificando as relações entre elas. Nos casos onde usuários e itens possuem características semelhantes, infere-se que o item é relevante ao usuário [Balabanović and Shoham 1997].

¹<http://dblp.uni-trier.de/>

O grande desafio dos sistemas de recomendação é identificar corretamente as necessidades dos usuários. Por isso, as redes sociais têm sido cada vez mais utilizadas no âmbito dos sistemas de recomendação, pois, através delas, é possível extrair diversos tipos de informação dos usuários e definir o perfil dos mesmos.

O assunto de rede sociais pode ser compreendido como o estudo das relações entre indivíduos e as implicações dessas relações [Wasserman 1994]. Os indivíduos são conhecidos como atores, esses atores podem representar organizações, conjuntos de pessoas ou um único indivíduo e estão conectados entre si através de laços relacionais.

As redes sociais podem ser definidas como um conjunto finito de atores, ou grupos de atores, e as relações definidas entre eles. Diferentes tipos de redes sociais podem ser observadas com finalidades diversas. O uso de redes sociais, em conjunto com os sistemas de recomendação, pode acrescentar informações na forma de relacionamentos entre os usuários. Em geral o relacionamento entre os usuários pode ser implícito ou explícito.

Neste trabalho foi utilizada uma rede social científica. Este tipo de rede social refere-se a redes nas quais os atores são representações de pesquisadores e as ligações são estabelecidas por relacionamentos de co-autoria entre os autores [Goldenberg et al. 2010, Ströele et al. 2013]. Essas redes podem ser estruturadas e representadas através de Grafos. Os Grafos são modelos versáteis que possibilitam a análise de diferentes problemas onde os pontos e suas conexões possuem alguma interpretação, seja ela física ou conceitual. Neste trabalho, os grafos bipartidos foram utilizados para modelar a rede social semântica utilizada como base para realizar recomendações.

4. Trabalhos Relacionados

Como verificado em [Shi et al. 2014] as abordagens baseadas em grafos vem sendo bem estudadas e muito utilizadas, principalmente na predição de links em redes sociais. Diversos trabalhos atacam o problema de recomendações a partir de uma abordagem que utiliza grafos bipartidos utilizado para diferentes propósitos. Em [Kunegis et al. 2010] o objetivo é especializar algoritmos de recomendação para o caso de grafos bipartidos. De acordo com os estudos, dentre os métodos de predição avaliados, os modelos de recomendação algébricos, que levam todo o grafo em consideração, podem ser aplicados.

Outros trabalhos utilizam a ideia de arestas com peso em grafos bipartidos, em [Sawant 2013] as arestas possuem peso referente à qualificação. Neste trabalho uma rede de alocação dos recursos foi utilizada para explicitar a similaridade entre pares de usuários e pares de objetos, que são utilizados para predições de recomendações e qualificações. Eles concluíram que as predições de recomendação são afetadas de maneira considerável pela acurácia da função de similaridade entre as entidades. Em [Ting et al. 2013] o conceito de peso nas arestas para demonstrar o interesse entre usuários e objetos também é utilizado. O objetivo é aumentar a viabilidade do uso de algoritmos para recomendação em grafos bipartidos.

A abordagem de redes semânticas também pode ser vista em [Maehara et al. 2012] onde o objetivo é recomendar itens de exibição em museus. A rede semântica é construída utilizando apenas as características dos itens, possibilitando os usuários a entender o contexto que engloba o item, despertando o interesse dos mesmos. Em [Shieh et al. 2008], uma abordagem colaborativa é escolhida para

recomendação de imagens a partir de um grafo semântico. O grafo semântico é construído a partir da redução de grafos bipartidos entre autores e termos. Nesta abordagem, o grafo bipartido é reduzido a um grafo unidimensional onde as arestas possuem peso e os nós representam os autores.

Diferentemente de [Kunegis et al. 2010, Shi et al. 2014] a abordagem deste trabalho utiliza grafos bipartidos com pesos em suas arestas para auxiliar o processo de recomendação. Em [Sawant 2013, Ting et al. 2013] também é adotado o peso nas arestas, mas esse peso não é definido com base nas informações semânticas dos autores e dos objetos.

Neste trabalho as arestas possuem um peso que representa o interesse que um autor possui pelo objeto com qual ele está conectado. Diferentemente de [Maehara et al. 2012], esse interesse é definido pela análise semântica tanto dos autores quanto dos objetos. Ao final tem-se um grafo semântico bipartido utilizado para recomendação de objetos aos autores, o que torna essa abordagem diferente de [Shieh et al. 2008].

5. Modelo de Recomendação baseado em Grafo Bipartido

Nesta seção serão apresentadas as etapas que compõem o modelo para resolução do problema de recomendação com filtragem baseada em conteúdo.

Na Figura 1 é apresentada uma visão geral do *workflow* que representa a solução proposta neste trabalho. No primeiro passo foi realizada a coleta dos dados nas fontes de informação selecionadas. Após a extração e tratamento dos dados, eles são transmitidos ao analisador de conteúdo que irá fazer a análise semântica dos mesmos.

Depois da análise do conteúdo ter sido realizada as informações semânticas são extraídas para que o grafo semântico seja construído pelo quantificador de relacionamentos. Esse módulo é responsável por dar pesos aos relacionamentos entre os autores e itens a partir das informações semânticas identificadas anteriormente. Com base no grafo obtido pela relação entre autores e itens é possível realizar as recomendações. De posse do grafo semântico o recomendador irá ordenar os itens, segundo os interesses do autor, e apresentar a lista de itens recomendados a ele. As próximas seções apresentam os detalhes do desenvolvimento de cada uma dessas etapas.

5.1. Fontes de Informação

As fontes de informação são repositórios que contém dados para extração de conteúdos. Esses repositórios devem possuir dados que contemplam as necessidades de informação de um usuário ou grupos de usuários. Neste trabalho serão usadas como fontes de informações a DBLP e o YouTube.

5.1.1. Base de Pesquisadores

A DBLP possui uma vasta coleção de dados com diversos tipos de informação relativas às publicações da área de ciência da computação, como artigos, *journals*, livros e outros. Os dados estão disponíveis no formato XML e foram mapeados para um banco de dados relacional. Esse mapeamento foi feito com auxílio da biblioteca *JAVA SAXParser* permitindo a análise das diferentes *tags* encontradas para realizar a identificação da estrutura do documento.

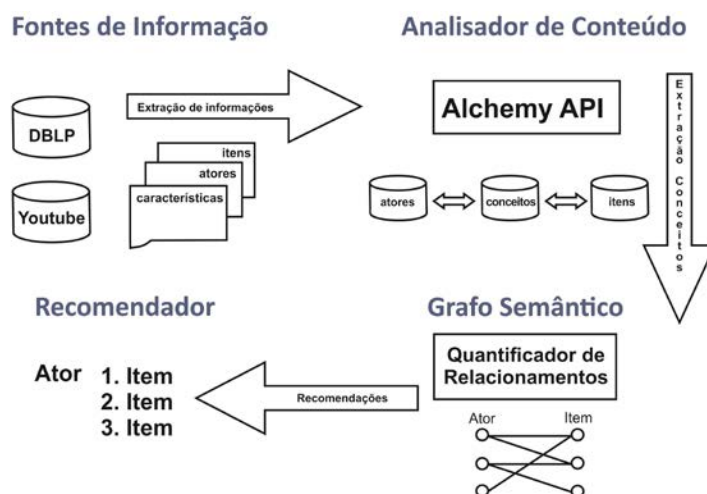


Figura 1. Workflow desenvolvido

Em uma etapa posterior ao processamento das informações do arquivo XML foi realizada a carga em um banco de dados *PostgreSQL*. Após a consolidação dos dados foi possível selecionar as informações dos autores e dos artigos publicados por eles para o módulo analisador de conteúdo.

5.1.2. Base de Vídeos

Conforme dito anteriormente, os objetos a serem recomendados pelo sistema de recomendação são os vídeos. Para isso, é necessário definir um repositório com esses objetos e, neste trabalho, será utilizado o YouTube. Esse repositório foi escolhido devido à grande quantidade e diversidade dos vídeos na plataforma. A variedade dos vídeos possibilita a recomendação para diferentes usuários em diferentes contextos.

O conteúdo do YouTube pode ser coletado a partir de consultas no site ou utilizando a *YouTube Data API*². Neste trabalho a coleta dos dados foi realizada, exclusivamente, através da API citada anteriormente. A linguagem JAVA foi utilizada para realizar as chamadas a API para capturar dados de vídeos, *playlists* e canais.

Com a API pronta para receber as requisições foi possível iniciar o processo de extração dos vídeos, as respostas da API são transmitidas no formato JSON e convertidas para classes JAVA. Para extração dos dados foi realizada uma chamada no método *search* retornando 50 resultados. Este método teve como parâmetro os conceitos identificados nos autores. A partir do retorno obtido foi possível realizar a coleta dos dados pertinentes de cada vídeo. Os dados coletados foram transcritos para o banco de dados *PostgreSQL*.

5.2. Análise do Conteúdo

Para a análise do conteúdo e extração de informações semânticas foi utilizada a *AlchemyAPI*³. A API realiza uma análise de textos, estruturando-os e tornando-os legíveis aos computadores.

²<https://developers.google.com/youtube/v3/>

³<http://www.alchemyapi.com/>

Essa API foi aplicada para a extração de informações semânticas relacionadas aos vídeos e pesquisadores. Ela realiza análises semânticas de texto, permitindo a extração de conjuntos de conceitos relevantes ao texto analisado. Os conceitos identificados são as ideias discutidas no texto, independente se elas são mencionadas explicitamente ou não. Somente os textos da língua inglesa foram avaliados pela API.

Além dos conceitos de cada vídeo e autor, um coeficiente de relevância é gerado por essa API indicando a importância de cada conceito para um dado vídeo ou autor. De acordo com estudos realizados por [Saif et al. 2012] foi observado que os resultados da extração de conceitos apresentados pela plataforma são bons quando comparados a outras soluções, fundamentando o uso neste trabalho.

Para os vídeos, a extração de conceitos ocorreu a partir da análise do título e descrição. A especificidade de um conceito para um dado vídeo também foi determinada pela *AlchemyAPI*. Desta forma, foram armazenados os pares compostos por conceitos e vídeos, assim como sua relevância.

A identificação dos conceitos associados aos autores ocorreu através da análise de todos os títulos dos trabalhos publicados por esses autores. Assim como foi feito para os vídeos, também foram armazenados os pares de conceitos e autores com suas respectivas relevâncias.

Após a identificação dos conceitos foram construídos dois grafos bipartidos. Ambos possuem dois tipos de nós, no primeiro estão representados os autores e os conceitos extraídos, enquanto no segundo grafo são representados os vídeos e os seus conceitos. A aresta dos grafos bipartidos gerados, neste momento, tem como peso o valor da relevância entre as suas extremidades, essa medida foi extraída pelo extrator de conteúdo e representa a importância do conceito aos dados analisados pela API.

5.3. Construção do Grafo Semântico

O grafo semântico foi construído a partir da união dos dois grafos bipartidos gerados a partir do módulo de *Análise do Conteúdo* (descrito anteriormente), conforme representado na Figura 2. Os nós do grafo semântico representam os autores (pesquisadores) e itens (vídeos) e as arestas que os relacionam possuem um peso que representa a relevância dos vídeos para os pesquisadores. O peso dessas relações foi calculado com a Equação 1.

$$R = \frac{\sum_{i \in C} (CP_i + CV_i)}{1 + \sum_{j \in CV} CV_j} \quad (1)$$

Onde CP e CV são os conjuntos das relevâncias dos conceitos extraídos dos pesquisadores e dos vídeos, respectivamente, e $C = \{CP \cap CV\}$ é o conjunto das relevâncias dos conceitos comuns aos pesquisadores e aos vídeos. Assim, R representa o quão relevante um vídeo é para determinado pesquisador.

Nesta equação, no numerador são somadas as relevâncias referentes aos conceitos presentes nos dois nós (pesquisadores e vídeos). O denominador é a soma de todos os conceitos que apenas o vídeo possui, este denominador tem como objetivo penalizar a relevância do relacionamento para os conceitos que não são comuns aos dois nós. Como as

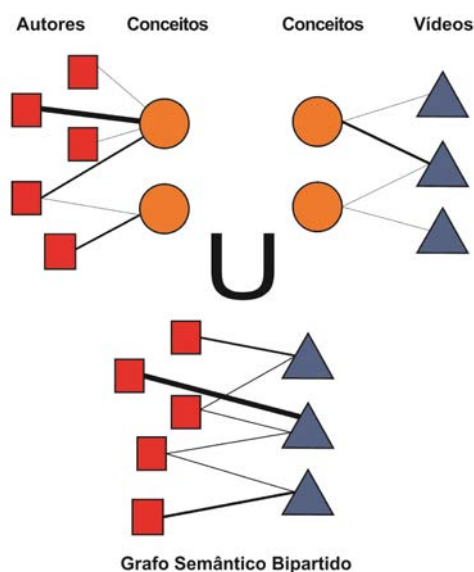


Figura 2. Representação do grafo semântico.

relevâncias fornecidas pela API variam entre 0 e 1, onde os valores próximos a 1 indicam maior importância, o intervalo da equação é $[0, 2n]$ onde n é a quantidade de conceitos comuns. Desta forma, os valores fornecidos pela equação que estão mais próximos a 0 indicam uma menor aderência entre os nós e valores mais distantes de 0 indicam maior aderência entre os nós.

Somente os conceitos presentes nos dois grafos gerados foram considerados. Desta forma somente os conceitos que aparecem nos dois grafos podem ser utilizados para ajustar a relevância dos relacionamentos.

5.4. Recomendação

Foi realizada uma análise semântica dos dados extraídos com a finalidade de representar o perfil dos atores e itens, que, neste trabalho, são pesquisadores e vídeos, respectivamente. A Equação 1 indica a aderência existente entre os pesquisadores e os vídeos que serão recomendados. Assim, listas de recomendação foram geradas analisando os valores das arestas do grafo semântico ordenadas de forma decrescente, visto que os maiores valores indicam a maior aderência entre um autor e um vídeo.

6. Resultados

Após o desenvolvimento e execução da solução proposta foram identificados os dois grafos gerados no processo descrito na Subseção 5.2. Foram avaliadas 89.973 instâncias com um total de 22.428 conceitos identificados gerando um total de 626.445 relacionamentos.

Para a construção do grafo semântico, nem todos os autores possuíam conceitos relacionados aos vídeos. Mesmo assim, a quantidade de autores e vídeos que compõem o grafo contemplou a maioria dos atores e itens classificados. Aproximadamente 78% dos autores e 72% dos vídeos estão presentes no grafo semântico. Esses nós se relacionaram através de 1598 conceitos.

A Tabela 1 mostra os conceitos que são compartilhados pelo maior número de autores e vídeos. Pode ser observado que a quantidade dos conceitos mais identificados

para os autores foi bem maior, isto aconteceu pois o conjunto de instâncias de autores analisados foi superior aos vídeos.

Tipo	Nome do conceito	Quantidade
Vídeo	Vector space	215
	Mathematics	202
	Group	144
Autor	Mathematics	8448
	Algorithm	8413
	Logic	4215

Tabela 1. Conceitos mais comuns.

Autor	Conceito	Relevância
22	Logic	0.778654
	Reasoning	0.719578
	Problem solving	0.689492

Tabela 2. Conceitos identificados para o autor 22.

A seguir seguem dados coletados de um autor presente no grafo semântico obtido através do *workflow* desenvolvido neste trabalho. A Tabela 2 mostra os conceitos identificados para o autor 22. Os conceitos apresentados são apenas aqueles em que existe algum vídeo com o mesmo conceito.

Na Tabela 3 podemos ver os conceitos associados a alguns vídeos que se relacionam com o autor 22. Com as informações referentes à relevância dos conceitos para os autores o grafo bipartido é construído. Da mesma forma, é construído um grafo bipartido que representa os vídeos e seus conceitos. Nos dois grafos construídos o peso das arestas representa a relevância encontrada pelo Analisador de Conteúdo.

A Tabela 4 representa a lista de recomendação para o autor 22. Essas listas de recomendação foram concebidas a partir da construção do grafo semântico, que representa a união dos dois grafos bipartidos construídos.

Vídeo	Conceito	Relevância
DIKkujAieTY	Logic	0.988
	Reasoning	0.638043
	Inductive reasoning	0.618041
	Fallacy	0.617589
	Deductive reasoning	0.608903
	Inference	0.596085
	Analogy	0.593086
	Abductive reasoning	0.55106
ervHbKa7R5g	Problem solving	0.92607
hq1bUM2tyg0	Problem solving	0.865606
kpqSeXpiM6k	Problem solving	0.895888
SpDIXJ2I2D4	Problem solving	0.954089
	Problem	0.470192
Zyq6TmQVBxk	Problem solving	0.895888

Tabela 3. Conceitos identificados para os vídeos relacionados ao autor 22.

Vídeo	Relevância
ervHbKa7R5g	1,615562
kpqSeXpiM6k	1,585380
Zyq6TmQVBxk	1,585380
hq1bUM2tyg0	1,555098
SpDIXJ2I2D4	1,117936
DIKkujAieTY	0,681447

Tabela 4. Listas de recomendação para o autor 22.

7. Avaliação

Este trabalho foi avaliado com base em duas provas de conceitos (PoC) desenvolvidas segundo o método GQM [Basili 1992]. As provas de conceito foram realizadas com

usuários reais, professores do departamento de ciência da computação. A avaliação foi feita através de um questionário onde esses professores avaliaram diferentes questões sobre a abordagem proposta neste trabalho.

7.1. Primeira PoC: definição de perfil por uma rede social científica

Nesta primeira prova de conceito é analisada a definição do perfil dos professores para avaliar a viabilidade de definição de perfil utilizando uma rede social científica. O objetivo é verificar se as informações extraídas estão corretas, levando em consideração a opinião do usuário.

Através dos conceitos extraídos para cada um dos cinco professores foi possível identificar o interesse de pesquisa deles. Dos cinco professores convidados a participar desta avaliação quatro deles desenvolvem seus trabalhos em linhas de pesquisa distintas, embora todos sejam da grande área da computação. Em média foram extraídos sete interesses de pesquisa (conceitos) de cada professor.

Após identificar esses interesses os professores foram convidados a avaliá-los respondendo à seguinte pergunta: *As informações extraídas da DBLP representam de fato seus interesses de pesquisa?*

Quando são analisadas as avaliações em (a) na Figura 3 verifica-se que quatro professores concordam que os interesses identificados pelo processo de extração são aderentes aos seus perfis de pesquisa. Nenhum professor discordou do perfil identificado tenha relação com seus interesses de pesquisa, mas houve um professor que indicou ser neutro em relação à esse questionamento.

Esta primeira prova de conceito mostra a viabilidade de extrair informações relacionadas aos interesses de pesquisa dos professores através de uma rede social científica. Têm-se indícios de que o uso das redes sociais científicas é viável no processo de definição dos interesses de pesquisa dos professores e que, tais interesses, são compatíveis com a realidade dos professores.

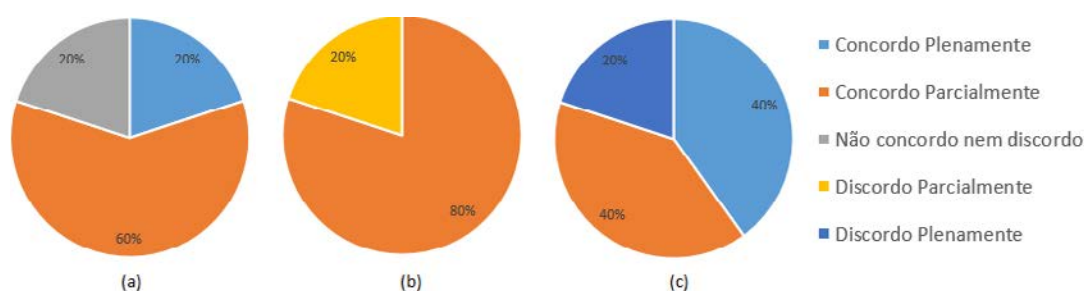


Figura 3. Gráficos de resposta das Perguntas (a), (b) e (c)

7.2. Segunda PoC: aderência das recomendações aos interesses dos professores

O objetivo da segunda prova de conceito é analisar a viabilidade da geração de recomendações e avaliar a qualidade das mesmas com base na aderência aos interesses de pesquisa dos professores considerando a opinião do usuário.

Para a realização desta prova de conceito foi solicitado ao usuário que ele avaliasse uma lista de vídeos recomendados a ele. Essa lista foi ordenada segundo a relevância

dos termos extraídos para cada autor, calculada de acordo com a Equação 1. Foram recomendados entre 7 e 10 vídeos para cada autor.

Os cinco professores foram convidados a avaliar as recomendações geradas para os conteúdos extraídos para cada um deles. Para realizar a avaliação eles responderam às seguintes perguntas: (b) O tema dos vídeos recomendados a você está coerente com as informações extraídas da DBLP? (c) Você compartilharia o vídeo recomendado para algum aluno ou professor da mesma área?

O gráfico (b) da Figura 3 apresenta as avaliações dos professores com relação a primeira pergunta. Pode-se observar que quatro, dentre os cinco professores, concordam parcialmente que as recomendações são coerentes com as informações extraídas da DBLP. Um professor discordou parcialmente que haja uma relação entre as recomendações e os conceitos extraídos.

Também foi solicitado aos professores que eles deixassem seus comentários e críticas para que o sistema de recomendação pudesse ser aprimorado. Nesses comentários foi possível compreender melhor as avaliações feitas por eles. Em geral, sempre há algum vídeo que não é condizente com os interesses de pesquisa dos professores e, por isso, nenhum deles concordou plenamente com as recomendações. O professor que discordou parcialmente recebeu recomendação de vídeos de um canal de humor que abordava, de alguma maneira, os conceitos extraídos da DBLP. Em um trecho dos seus comentários o professor diz que:

“...acredito que possam incluir algum tipo de filtro para selecionar canais relacionados a educação/pesquisa, já que o foco é identificar vídeos associados ao interesse de pesquisa do indivíduo.”

O gráfico (c) da Figura 3 consolida as respostas dos professores à segunda pergunta. O objetivo dessa pergunta é identificar se existe alguma recomendação que o professor considera relevante ao ponto de repassar o vídeo sugerido. Observa-se que quatro professores consideram os vídeos interessantes e recomendariam os mesmos para outras pessoas. Um professor, o mesmo que discordou parcialmente na questão anterior, discorda plenamente, ou seja, ele não recomendaria os vídeos a outras pessoas. Este professor considerou os vídeos superficiais aos seus temas de pesquisa e, por isso, não recomendaria. Os resultados da primeira prova de conceito indicam que a estratégia de definição do perfil de pesquisadores a partir de uma rede social científica foi bem avaliada pelos professores. Dentre os professores que avaliaram essa estratégia apenas um não concordou com o perfil definido para ele.

Da mesma forma, os resultados da segunda prova de conceito também trouxeram indicativos de que as recomendações feitas aos professores foram, de certa forma, coerentes aos perfis identificados. Como era esperado, o professor que não concordou com a definição de seu perfil também não concordou com os vídeos recomendados.

Assim sendo, os resultados observados mostram a viabilidade das estratégias de extração de características em redes sociais científicas e de recomendação de vídeos.

7.3. Ameaças a validade

Existem algumas ameaças à validade deste estudo. O estudo simula que os professores estavam precisando dos vídeos que foram recomendados. Portanto, é importante que

esse estudo seja expandido para que os professores possam acessá-lo quando realmente estiverem necessitando de alguma recomendação.

A qualidade dos termos extraídos para identificar o perfil dos professores também é uma ameaça a validade. Há um risco de que, quando outros professores forem selecionados, termos muito genéricos ou ambíguos sejam identificados, impactando diretamente na qualidade das recomendações.

Outra ameaça é a quantidade de professores convidados a fazerem a avaliação das estratégias apresentadas neste trabalho. Embora tenham sido selecionados professores de diferentes áreas de pesquisa, sabe-se que é necessário avaliar a proposta com um maior quantitativo de pessoas.

8. Considerações Finais

O objetivo deste trabalho foi desenvolver um modelo de recomendação baseado em informações semânticas e, a partir do modelo proposto, gerar recomendações. Para isso, foi realizada uma pesquisa bibliográfica a fim de fundamentar a abordagem da solução proposta e identificar os trabalhos relacionados.

Após a modelagem, foi realizada a extração dos dados e, de posse dos dados, foram extraídos conceitos com o objetivo de descrever os itens e atores utilizados no sistema de recomendação. De posse dos atores e itens, com seus respectivos perfis, foi possível quantificar a relação entre eles e criar um grafo semântico bipartido. A partir do grafo semântico construído foi possível extrair uma lista de recomendações ordenada pela relevância entre os nós.

Com o intuito de avaliar a viabilidade da solução proposta foram realizadas duas provas de conceitos. Um questionário foi desenvolvido e cinco professores o responderam. Com base nas respostas obtidas verificou-se que o modelo de recomendação baseado em grafos semânticos é viável. Além disso, avaliando as sugestões desses professores, algumas evoluções podem ser aplicadas ao modelo para aprimorar as recomendações.

Como trabalhos futuros tem-se a modificação do modelo proposto adicionando o *feedback* fornecido pelo usuário após a recomendação para auxiliar as próximas recomendações. A partir do *feedback* pode ser utilizada uma abordagem de filtragem híbrida para fazer recomendações utilizando, além das informações semânticas, a opinião dos usuários com perfil semelhantes. Outra oportunidade para trabalhos futuros é abordar o uso de ontologias na etapa de análise do conteúdo afim de agrupar os conceitos dentro de um domínio, podendo assim modificar a qualificação entre os relacionamentos e alterar a lista de recomendação.

9. Agradecimentos

Agradecemos a CAPES, CNPQ, FAPEMIG e UFJF.

Referências

- [Adomavicius and Tuzhilin 2005] Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749.

- [Baeza-Yates et al. 1999] Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- [Balabanović and Shoham 1997] Balabanović, M. and Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72.
- [Basili 1992] Basili, V. R. (1992). Software modeling and measurement: the goal/question/metric paradigm.
- [Goldenberg et al. 2010] Goldenberg, J., Libai, B., Muller, E., and Stremersch, S. (2010). Database submission-the evolving social network of marketing scholars. *Marketing Science*, 29(3):561–567.
- [Han and Kamber 2006] Han, J. and Kamber, M. (2006). *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann.
- [Kunegis et al. 2010] Kunegis, J., De Luca, E. W., and Albayrak, S. (2010). The link prediction problem in bipartite networks. In *Computational intelligence for knowledge-based systems design*, pages 380–389. Springer.
- [Maehara et al. 2012] Maehara, C., Yatsugi, K., Kim, D., and Ushiyama, T. (2012). An exhibit recommendation system based on semantic networks for museum. In *Innovations in Intelligent Machines-2*, pages 131–141. Springer.
- [Oestreicher-Singer and Sundararajan 2010] Oestreicher-Singer, G. and Sundararajan, A. (2010). Recommendation networks and the long tail of electronic commerce. Available at SSRN 1324064.
- [Saif et al. 2012] Saif, H., He, Y., and Alani, H. (2012). Semantic sentiment analysis of twitter. In *The Semantic Web-ISWC 2012*, pages 508–524. Springer.
- [Sawant 2013] Sawant, S. (2013). Collaborative filtering using weighted bipartite graph projection: a recommendation system for yelp. *CS224W: Social and Information Network Analysis (December 10, 2013)*.
- [Shi et al. 2014] Shi, Y., Larson, M., and Hanjalic, A. (2014). Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, 47(1):3.
- [Shieh et al. 2008] Shieh, J.-R., Yeh, Y.-T., Lin, C.-H., Lin, C.-Y., and Wu, J.-L. (2008). Collaborative knowledge semantic graph image search. In *Proceedings of the 17th international conference on World Wide Web*, pages 1055–1056. ACM.
- [Ströele et al. 2013] Ströele, V., Zimbrão, G., and Souza, J. M. (2013). Group and link analysis of multi-relational scientific social networks. *The Journal of Systems and Software*, 86:1819–1830.
- [Ting et al. 2013] Ting, Y., Yan, C., and Xiang-wei, M. (2013). Personalized recommendation system based on web log mining and weighted bipartite graph. In *Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on*, pages 587–590. IEEE.
- [Wasserman 1994] Wasserman, S. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.