



Arthur Silva Morato

Sistema de Recomendação Utilizando a Recomendação por Conteúdo

Mato Grosso do Sul - Brasil

27 de setembro de 2014

Trabalho de Conclusão de Curso

Arthur Silva Morato

Sistema de Recomendação Utilizando a Recomendação por Conteúdo

Dissertação apresentada ao curso de graduação em Ciência da Computação da Universidade Federal de Mato Grosso do Sul como trabalho de conclusão de curso: requisito obrigatório para colação de grau e obtenção do título de Bacharel em Ciência da Computação.

Fundação Universidade Federal de Mato Grosso do Sul

Campus de Ponta Porã – MS

Ciência da Computação – Bacharelado

Orientador: Me. Daniel Matte Freitas

Mato Grosso do Sul - Brasil

27 de setembro de 2014

Me. Daniel Matte Freitas

AGRADECIMENTOS

*A Deus pelas bênçãos sem fim, e mais um monte de blábláblá que tenho que
bolar depois!*

*A meus pais por tudo, em especial: pela grana que me manteu todo esse tempo
longe de casa.*

A meus amigos e colegas da xuera sem fim.

A meus professores amados

A meu orientador.

Dedico a mim e só a mim! Vlw flw. Ass. Irmauxinho

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

EXP *Exemplo*

LISTA DE SÍMBOLOS

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
\in	Pertence

SUMÁRIO

Agradecimentos	3
Resumo	15
Abstract	15
1 INTRODUÇÃO	19
1.1 Justificativa	20
1.2 Objetivos	21
2 SISTEMA DE RECOMENDAÇÃO	23
2.1 Exemplos de sistemas que usam recomendação	23
2.1.1 Subseção	24
3 O PROBLEMA	25
4 CONSIDERAÇÕES FINAIS	27
Referências	29

RESUMO

Este trabalho tem como objetivo apresentar as principais técnicas sobre os Sistemas de Recomendação de conteúdo, entre muitas técnicas conhecidas, que foram derivadas ou baseadas nessas técnicas ou em outras, apresentando de forma clara todos os seus mecanismos de funcionamento e seus respectivos resultados encontrados em trabalhos realizados anteriormente, proporcionando uma prévia dos dados obtidos em pesquisas e estudos realizados desde os primórdios do estudo sobre a recomendação e das primeiras hipóteses de recomendação por um sistema independente, onde se utiliza dados coletados para poder prever o que poderia ser mais relevante, sendo assim, com uma maior probabilidade de acerto próximo passo, para determinado fim baseado em alguma métrica estabelecida pelo sistema, aprimorando todo o conjunto do sistema onde foi aplicado a recomendação. Para demonstração de desempenho e visualização dos resultados obtidos, será proposto neste trabalho o desenvolvimento de um framework para recomendação generalizada, onde o algoritmo passará por um processo de aprendizagem de máquina indutivo, utilizando uma base de dados que será subdividida em conjunto de teste e conjunto de aprendizagem, para poder generalizar novos casos e classificar as informações do conjunto de teste, utilizando o “aprendizado” obtido por meio do conjunto de aprendizagem.

ABSTRACT

This work aims to present the main techniques on Recommender Systems, among many known techniques, which were derived from or based on such technical or other, showing clearly all its operating mechanisms and their results found in studies previously conducted providing a preview of the data obtained in surveys and studies conducted since the beginning of the study and on the recommendation of the first hypotheses recommendation by an independent system, which uses the collected data in order to predict what could be more relevant, therefore, to a higher probability of success next step for a particular purpose based on some metric established by the system, improving the whole system where the recommendation was applied. To demonstrate performance and displaying the measurement results, it is proposed in this work to develop a generalized framework for recommendation where the algorithm will go through a learning process of induction machine using a database to be divided into the test set, and learning to generalize to new cases and sort the information of the test, using the 'learning' obtained through the learning set.

1 INTRODUÇÃO

A recomendação está em presente em muitos lugares e situações, uma vez que é preciso, na maioria da vezes, algum conhecimento sobre algo novo para se interessar por ele ao ponto do querer mais informações para poder confirmar o interesse, mas o primeiro obstáculo à ser derrubado é como saber o que terá nesse algo novo que poderá ser do interesse de quem ou o que, está recebendo a recomendação. Sem fazer distinção do que será recomendado, é natural o conhecimento de que todas as informações sobre um produto ou serviço é do interesse de quem está procurando, mas poder apresentar resultados baseados em suas preferências pode ser a principal etapa para definir se há realmente o interesse, ou não, proporcionando uma otimização significativa em todo o conjunto do sistema, que será refletido nos resultados que serão obtidos desse sistema. Entre pessoas, por exemplo, a preferência é ver se alguém já testou algum novo produto ou serviço, e qual foi a reação que isso proporcionol para a pessoa, para depois analisar se o novo produto pode, ou não, see bom, ou ruim, para si próprio.[1],[2].

Uma grande conquista para a tecnologia é a grande quantidade de conteúdo que pode ser compartilhado através de diversos meios de comunicação e de tráfego de dados, possibilitando a propagação de informação, acessível a todas as pessoas com acesso a tecnologia, otimizando as atividades em todos os âmbitos pessoais e tecnológicas. Um grande avanço se mostra no compartilhamento em massa de informações em todas as áreas de interesses, como notícias sobre esportes, tempo, política, ou sobre estudos e pesquisas, como conseguir encontrar áreas de pesquisa e estudo que estão em destaque nos dias atuais, ou sobre alimentação, como qual restaurante poderia ser o melhor para visitar em determinada cidade, também é demonstrado avanço sobre o compartilhamento de mídias digitais, como música, áudio, filmes, texto, etc. [5]. Alguns exemplos de sistemas que conseguem recomendar CDs, livros e uma outra infinidade de produtos pelo loja online Amazon.com [3], ou filmes e suas classificações pelo site MovieLens.com [6], ou livros e críticas de livros no site barnesandnoble.com, ou artigos acadêmicos, sites, lugares para conhecer ou visitar, culinária, produtos em geral, entre outros, procurando sempre o que for melhor para oferecer aos seus usuários, que estão buscando o que for mais interessante para ele. [4], [5], [6], [7].

Descobrir o mecanismo e as funcionalidades dos algoritmos de recomendação é descobrir como a recomendação desses sistemas, citados acima, conseguem oferecer um novo produto com uma grande taxa probabilidade de que um futuro novo usuário possa gostar, e continuar recomendando com cada vez mais precisão, utilizando como parâmetro o perfil de seus usuários. Avaliando essa tarefa de recomendação como um desafio, então

conseguir entender o funcionamento humano de recomendação e transformá-lo em dados através de um algoritmo que possa “aprender” a recomendar com a maior precisão possível, é uma provável solução para esse desafio. [2].

Neste trabalho, para uma melhor abordagem do assunto, será feita uma ampla abordagem sobre os tipos de sistemas de recomendação, alguns problemas na recomendação e suas soluções. Também é proposto neste documento, o desenvolvimento de um Framework para exemplificar o funcionamento de um algoritmo de recomendação, com a finalidade de visualizar os dados obtidos e verificar e comprovar a eficiência da recomendação por sistemas de Inteligência Artificial. Os resultados se encontrarão ao final desse documento.

Nas seções e subseções deste trabalho, será abordado o histórico da recomendação, exemplifica algumas técnicas mais utilizadas e seu funcionamento, e a elaboração, desenvolvimento e a avaliação dos resultados obtidos, como forma de visualização do desempenho do recomendador, as considerações finais e as respectivas referências bibliográficas.

1.1 JUSTIFICATIVA

É possível encontrar com facilidade resultados que são demonstrado em diversos trabalhos publicados e disponibilizados na internet, provando a eficiência do sistema de recomendação aplicado em diversas aplicações. A derivação desse trabalho e suas pesquisas dão seguimento ao raciocínio lógico de otimização encontrado na maior parte dos trabalhos referentes a recomendação.

Muitas vezes é preciso fazer escolhas sem ter experiência suficiente sobre as escolhas que nos são dispostas. Em nossa vida diária, buscamos por recomendação de outras pessoas que já tiveram experiências sobre o que você está procurando ou até mesmo por boatos feitos de boca em boca para descobrir quais as conclusões encontradas pelos outros, como pedir uma carta de recomendação, avaliações e comentários sobre livros e filmes que são publicadas em jornais, ou levantamentos gerais como encontrar um restaurante que lhe interesse pelo guia de restaurantes Zagat's. (??)

Na maioria das aplicações, a recomendação é feita para oferecer ao seu usuário um produto que ele tem uma chance maior de interesse por parte do consumidor, facilitando a procura desse usuário, se for o caso, e aprimorando a eficiência do serviço prestado pelo ofertante.

As empresas precisam estar preparadas para, no mínimo, oferecer produtos com diversas características para conseguir oferecer a característica que mais se adequa a determinado usuário. A tendência é de que as empresas de comércio eletrônico possa oferecer mais produtos relacionados com o gosto de todos na internet. No entanto, ao disponibilizar diversas características de seus produtos para seus usuários, torna-se complicado a busca do usuário pelo produto que ele está procurando. Uma solução

para esse problema de *overload* de informação é a utilização de Sistema de Recomendação. (??, 4.0).

A recomendação é provada como ferramenta fundamental em um sistema corporativo quando é possível encontrar empresas oferecendo prêmios para melhorar seus sistema de recomendação ou oferecer um novo que supere o seu já utilizado, como é o caso da Netflix.

Em Outubro de 2006, a empresa Netflix lançou um grande desafio com uma grande base de dados, contendo avaliações sobre filmes, e desafiou todo o setor de Ciência da Computação, responsável pela mineração de dados e aprendizagem de máquina, para desenvolver um sistema para calcular o mais próximo possível da precisão dos dados referentes aos filmes sobre a grande base de informações. Os melhores deveriam documentar todo o trabalho e estudo realizado para a solução do sistema e divulgar, de forma clara, como foi que conseguiu encontrar seus resultados. (??, 4.0).

A procura pelo algoritmo de recomendação perfeito fez com que muitas pesquisas se voltassem para essa área, formando uma globalização de informações e resultados, o que torna mais viável as pesquisas sobre recomendação em diversos ambitos sociais e tecnológicos.

1.2 OBJETIVOS

Com esse trabalho em desenvolvimento, os objetivos são de demonstrar o uso e a aplicação de um algoritmo de recomendação de conteúdo sobre uma base de dados disponível na internet, para descrever as situações com que os usuários podem se deparar em um situação comum.

O algoritmo de recomendação será submetido a um conjunto de dados obtidos pela internet, coletados por um site de agrupamento de informações sobre filmes, e ao receber os dados referentes ao histórico do usuário, o algoritmo coletará informações sobre suas preferências no banco de dados, correlacionando suas informações com as de outros usuário que compraram ou avaliaram produtos relacionados, buscando quais foram suas outras compras que tem atributos os recursos em comum com o que o usuário está pesquisando, e mostrar, de forma ordenada por relevância, ao usuário os produtos que mais tiveram relação entre as escolhas dos usuário, de acordo com o banco de dados.

Com a pesquisa feita no banco de dados, os resultados esperados são de que, mesmo com um grande número de informações sobre os dados do sistema, as informações que são armazenadas no banco de dados são alimentadas de forma a aprimorar mais o sistema de recomendação, aumentando a taxa de acerto na recomendação de futuros produtos.

2 SISTEMA DE RECOMENDAÇÃO

O objetivo dos Sistemas de Recomendação é, de forma geral, garantir o significado de recomendação para um ambiente formado por um grande grupo de pessoas interessadas em algo novo, independente do que seja, com a finalidade de oferecer algo com mais precisão e certeza. [8] Ele foi emergido através de importantes pesquisas independentes, na área de pesquisa tecnológicas, com seus primeiros dados de avanços e descobertas publicados em meados dos anos 90. [5] Desta época até os dias atuais, as pesquisas sobre recomendação subiram drasticamente, buscando o aprimoramento das técnicas e a expansão do conhecimento. [7] De acordo com [7], alguns fatos marcaram o grande crescimento das pesquisas e utilização dos sistemas de recomendação, como o uso ascendente das grandes empresas da internet, que utilizam a recomendação para agradar novos usuários e manter seus usuários, com a recomendação de conteúdos contidos no site que possam ser do interesse desse usuário. As grandes empresas que utilizam esse sistema hoje em dia são, NetFlix, [10] YouTube, Yahoo, Tripadvisor, Last.fm e IMDb. Muitas empresas estão passando a utilizar a recomendação para otimizar seu desempenho no mercado, e todas as empresas que já utilizam a recomendação buscam á sempre ter seus sistemas totalmente atualizados, para garantir a melhor eficiência em recomendação, como, por exemplo, o NetFlix, que se dispõe a pagar um milhão de dollar para a pessoa, ou grupo de pessoas, que conseguir desenvolver um sistema melhor que o já implementado deles. Isso mostra que o setor de recomendação da empresa representa uma parcela muito grande de seu retorno financeiro. Hoje existe uma organização promotora dedicada a conferências e workshops sobre sistema de recomendação que é a ACM Recommender Systems (RecSys), foi estabilizada em 2007 e agora é conhecida por realizar anualmente uma conferência sobre recomendação e suas tecnologias sobre pesquisas e aplicações. Universidades do mundo todos realizam trabalhos sobre sistema de recomendação. Mais hoje do que nunca. Recentemente, uma faculdade, ondesse fez famosa por possuir grandes conferências sobre ciência, lançou um livro sobre as técnicas de recomendação.

2.1 EXEMPLOS DE SISTEMAS QUE USAM RECOMENDAÇÃO

Sistemas de recomendação são encontrados em uso por vários sistemas de lojas de eCommerce [3], [4], [6], com a tarefa de sugerir aos seus usuários produtos que possam lhe interessar. Os produtos podem ser recomendados partindo de várias premissas, como, por exemplo, sendo o produto mais vendido de uma determinada seção em que o usuário esta visitando, ou o mais vendido de todo o site, ou baseado em dados coletados do usuário como referência, ou analisando compras feitas pelo usuário com os dados dos produtos

comprados no passado, correlacionando com produtos vizinhos que tenham atributos idênticos ou semelhantes, que seria utilizado como uma previsão de uma futura compra desses produtos vizinhos. [4] Uma tarefa manual, está sendo substituída por sistemas de recomendação, automatizando todo o processo de recomendação.[9]

Segundo [4], alguns sites usam uma ou mais variações de sistemas de recomendação em seus códigos, estando propensos a mudanças com o passar dos anos, para otimização e manutenção. Esses exemplos, em ordem alfabética, são:

- **Amazon.com** – Foi usando a seção de livros especificamente. Primeira recomendação é feita por ordem de livros mais comprados por usuários que também compraram o livro selecionado. A segunda recomendação é feita por trabalhos desenvolvidos pelo autor do livro que também foi comprado por outros usuários.
- **Cdnow.com** – A recomendação é feita de duas formas. A primeira é feita através da escolha de determinado álbum de algum CD, na página do álbum outros 10 álbuns são recomendados, que são relacionados ao álbum selecionado primeiro. A segunda recomendação é feita após visitar a página de determinado autor/banda, que recomenda aos seus usuários até 10 álbuns feitos pelo artista.
- **Ebay** – Em trabalho conjunto entre comprador e vendedor, o sistema do eBay cria um Feedback das atividades de outros usuários, com quem fizeram negócios, para montar a predição de produtos de interesse do usuário, que estão sendo vendidos pelo vendedor que ele já fez negócio.
- **Levis** – Você precisa se identificar informando qual é o seu sexo, e mais alguns atributos precisam ser preenchidos, como o tipo de roupa que está procurando e é montado um ranque indo de “larga isto” até “amo isto”.
- **Moviefinder.com** – A recomendação é feita por filmes que tenham a maior quantidade de atributos parecidos com o que foi alugado pelo usuário, e com a melhor avaliação feita por usuários que já assistiram o filmes.
- **Reel.com** – Uma ligação entre os filmes semelhantes é feita e anunciada quando o usuário estiver decidindo sobre qual filme ele vai escolher. Uma frase do tipo “Em NOMEDOFILMENOVO existe as seguintes semelhanças com NOMEDOFILMEANTIGO,...” onde “NOMEDOFILMEANTIGO” seria algum filme já assistido pelo usuário com a maior porcentagem de aproximação de atributos com o filme novo.

2.1.1 SUBSEÇÃO

3 O PROBLEMA

4 CONSIDERAÇÕES FINAIS

Just-in-time

REFERÊNCIAS

Nenhuma citação no texto.

James Bennett and Stan Lanning. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35, 2007. Citado na página 21.

Paul Resnick and Hal R Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997. Citado na página 20.

J Ben Schafer, Joseph Konstan, and John Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166. ACM, 1999. Citado na página 21.