

Sistemas de Recomendação - Comércio electrónico, hotelaria e restauração

Universidade do Minho, Braga, Portugal

Abstract. Os sistemas de recomendação surgiram devido à necessidade de diminuir a sobrecarga de informação a que os utilizadores estão sujeitos durante os seus processos de pesquisa. A informação apresentada em sistemas de recomendação pode ser determinada como informação semelhante proveniente de pesquisas realizadas no passado pelo utilizador, informação obtida através de utilizadores com preferências similares, e/ou que eventualmente já foi classificada como relevante, ou ambos. Neste documento estão presentes técnicas e algoritmos utilizados por sistemas de recomendação. Existem assim, variados algoritmos de *machine learning* que permitem o suporte destas técnicas, estando alguns latentes no presente artigo. Muitas empresas na área do comércio electrónico, hotelaria e restauração utilizam estes sistemas, não só para proporcionar uma experiência melhor ao consumidor, mas também para retirar as diversas vantagens que deles advêm.

Keywords: Recommender systems · Collaborative filtering · K-Nearest Neighbors · Content-based · Hybrid Recommendation · Knowledge-based · Demographic Recommendation · Decision Trees · Naive Bayes classifier · Support Vector Machine · Artificial Neural Networks.

1 Introdução

Nos dias de hoje, é muito provável que a maioria dos utilizadores da internet já se tenham deparado com sistemas de recomendação. Quando se procura por um produto numa plataforma de retalho *online*, possivelmente, irá aparecer, numa área dessa página, uma série de produtos adicionais que poderão ser do interesse do utilizador. Com o uso frequente da mesma plataforma, uma lista personalizada de recomendações aparecerá automaticamente assim que o utilizador entra. O sistema que determina que itens deverão ser exibidos para o visitante em particular é designado como sistema de recomendação. Posto isto, Ricci *et al.* [1] define que "*Recommender Systems (RSs) are software tools and techniques that provide suggestions for items that are most likely of interest to a particular user.*"

Em 1970 na Duke University, os sistemas de recomendação evoluíram como uma área de pesquisa independente, permitindo a introdução de diversas abordagens. O primeiro SR a surgir foi o Tapestry1, criado pelo centro de pesquisa Xerox Palo Alto. A razão que levou ao seu desenvolvimento foi a crescente quantidade de *e-mails* recebidos, principalmente os desnecessários [11].

Os principais alvos dos SRs são utilizadores que não têm informação ou experiência necessária para avaliar uma infinidade de opções disponíveis que uma

plataforma *online* possa oferecer. Rapidamente estes sistemas tornaram-se bastante populares sendo utilizados por sites conhecidos como Amazon, Pandora, Netflix, YouTube, Yahoo ou Tripadvisor [11]. Assim, o principal objetivo dos SRs é aumentar as vendas de produtos e, conseqüentemente, o lucro, daí a sua utilização por parte dos comerciantes.

2 Técnicas utilizadas por sistemas de recomendação

2.1 *Collaborative Filtering*

A filtragem colaborativa é uma das técnicas mais populares e mais implementadas nos sistemas de recomendação [2]. A ideia passa por recomendar ao utilizador itens, que outros utilizadores com gostos semelhantes consideraram interessantes no passado. Nesta abordagem, a proximidade nas preferências de dois utilizadores é calculada com base na semelhança entre o histórico de classificações que os utilizadores atribuíram a cada item.

Esta técnica está dividida em duas categorias: uma baseada em modelos e outra baseada em memória (também conhecida por vizinhança ou heurística) [1]. A primeira constrói padrões das diferentes iterações utilizador/item para gerar previsões automáticas. Utiliza a informação da iteração parcial para modelar um conjunto de parâmetros para os utilizadores e itens, que juntos podem reconstruir previsões probabilísticas sobre as iterações em falta [9]. Os algoritmos nesta categoria têm uma abordagem probabilística e percebem o processo de filtragem colaborativa como o cálculo do valor esperado da previsão de um utilizador, dado os valores das suas avaliações sobre outros itens. A segunda baseia-se na similaridade entre utilizadores na escolha dos itens ou na semelhança dos próprios itens. Na recomendação baseada em itens é calculada a similaridade entre os itens, associando a cada um deles um conjunto de vizinhos mais próximos. Por outro lado, na recomendação baseada nos utilizadores, é escolhido o subconjunto de utilizadores com maior valor de similaridade com o utilizador ativo, e é usada uma combinação ponderada das suas avaliações para calcular previsões para este utilizador [2].

2.2 *Content-based recommendation*

Nesta técnica, a recomendação é feita tendo por base o perfil do utilizador, usando características do conteúdo dos elementos que este avaliou no passado, sendo os avaliados positivamente os recomendados. Estas técnicas têm por base análise estatística ou algoritmos de *machine learning*, tendo em vista recomendações que façam sentido ao utilizador. Na maioria dos casos, os atributos dos itens são palavras-chave simples que são extraídas a partir da descrição dos mesmos [5]. Desta forma, por exemplo, se um utilizador tem a possibilidade de avaliar um restaurante com determinadas particularidades, então o sistema pode aprender a recomendar restaurantes com as mesmas características.

Esta técnica possui benefícios no que diz respeito ao facto de apenas se basear no perfil de um utilizador. Assim, se as preferências deste utilizador mudarem, esta técnica tem o potencial de reajustar a recomendação num curto espaço de tempo.

No entanto, nem tudo são benefícios. Esta técnica depende dos meta-dados dos itens, existindo assim, a necessidade de ter descrições ricas dos itens bem como uma descrição organizada do perfil do utilizador, antes da recomendação ser feita a este. Desta forma, são necessários dados descritivos de qualidade para que a precisão do sistema de recomendação não seja posta em causa. A especialização excessiva de conteúdo é outro problema importante desta técnica, uma vez que os utilizadores estão restritos a receber recomendações semelhantes aos itens já definidos nos seus perfis.

2.3 *Hybrid Recommendation*

A elaboração de sistemas de recomendação híbridos combina os pontos fortes de diferentes algoritmos e modelos de forma a superar algumas das lacunas e problemas que estes possam possuir. De realçar que o facto de associar diferentes algoritmos faz com se torne mais forte e menos vulnerável a eventuais falhas e carências [4]. Existem vários métodos, tais como o *meta-level*, o *switching hybridization* e o *mixed hybridization*. No caso do *meta-level*, o modelo gerado por uma técnica de recomendação é usado como entrada para outra, tornando o modelo gerado mais rico em informações [5].

Outro exemplo são os sistemas de *switching hybridization*, onde existem trocas de técnicas, isto é, alternarem entre diversos métodos de forma a tornar a experiência do utilizador mais precisa. Como consequência, esta aumenta a complexidade do processo de recomendação [5].

Existem também os *mixed hybridization* que são caracterizados como sistemas mistos que combinam resultados de recomendações de diferentes técnicas de recomendação em simultâneo. Assim, cada item tem várias recomendações associadas de diferentes técnicas de recomendação, tornando o sistema mais confiável na informação disponibilizada [2].

2.4 *Knowledge-based Recommendation*

Há inúmeros casos em que os SRs *collaborative filtering* e *content-based* não se adequam, pois os utilizadores querem definir explicitamente os seus requisitos, por exemplo, na escolha do carro. Nestes casos é utilizado o sistema de recomendação *knowledge-based*. No entanto, esta abordagem está dependente da aquisição de conhecimento, ou seja, os utilizadores especificam os requisitos e o sistema tenta encontrar uma solução. Caso não consiga encontrar uma solução, o utilizador deve mudar os requisitos [2].

Esta teoria está dividida em dois tipos, *case-based* e *constrait-based*, em que ambos recolhem os requisitos dos utilizadores e apresentam uma solução. Estes princípios diferem na forma como usam o conhecimento adquirido, ou seja, como calculam a solução que apresentam.

Nos sistemas *case-based* são usadas métricas de similaridade para identificar itens semelhantes aos requisitos específicos dos clientes. Nesta abordagem a função

de similaridade estima quanto as necessidades do utilizador (descrição) correspondem à solução do problema (recomendações). Enquanto nos sistemas *constrained-based* os itens recomendados são obtidos com base num conjunto de regras sobre a relação entre os requisitos do cliente e os recursos do item [1].

2.5 Demographic Recommendation

Este tipo de sistema recomenda itens baseando-se no perfil demográfico do utilizador. Pressupõe-se que sejam criadas recomendações distintas para nichos demográficos diferentes. Muitos *sites* adotam soluções de personalização simples e eficazes com base na demografia. Por exemplo, com base no seu idioma ou país, os utilizadores são enviados para *sites* específicos. Outra forma é personalizar sugestões de acordo com a idade do utilizador [1].

Embora as abordagens citadas tenham sido bastante populares em áreas como a do marketing, houve relativamente pouca pesquisa apropriada sobre SRs em sistemas demográficos [1]. Para além disso, estes sistemas geralmente não fornecem os melhores resultados de forma independente, estes aumentam de modo significativo o poder de outros sistemas de recomendação como componente de modelos híbridos ou de conjuntos. Às vezes, técnicas demográficas são combinadas com SRs baseados no conhecimento para aumentar a sua robustez [3].

3 Vantagens

3.1 Para o promotor da recomendação

Qualquer entidade de negócio possui vários objetivos, que passam por aumentar a receita, os clientes e melhorar a eficiência. Neste sentido, os SRs auxiliam a que estes objetivos sejam alcançados, possuindo, assim, um conjunto de vantagens para a entidade que o promove, descritas em seguida.

- **Direcionar o tráfego:** Para além dos SRs aumentarem o tráfego nos *sites*, a partir de possíveis *emails* ou mensagens que sejam enviadas a clientes, permitem direcionar o tráfego a clientes habituais e categorizados por estes.
- **Maior controlo de marketing e inventário:** Permite perceber quais os produtos/serviços mais comprados/utilizados, para a entidade promotora reforçar as suas estratégias de marketing. Para além disto, permite controlar o inventário no que diz respeito a esses mesmos produtos.
- **Aumento do envolvimento com os clientes:** Sendo feitas recomendações úteis aos clientes, de acordo com os seus gostos, o aumento do envolvimento destes é maior. Isto permite que se sintam mais relacionados com o promotor, passando de comuns compradores a efetivos clientes. Pela ideia construída de que o sistema reconhece as suas preferências, os clientes estarão satisfeitos, levando à sua retenção.
- **Aumento das vendas:** Uma vez que os sistemas recomendam aos utilizadores produtos/serviços que vão de encontro às suas preferências e inter-relacionados, estes irão adquirir mais produtos de cada vez que efetuarem uma compra. Isto levará a um conseqüente aumento das vendas generalizado, sendo o comércio eletrónico o que melhor reflete esta vantagem.

3.2 Para o alvo da recomendação

No que diz respeito ao alvo de recomendação, estes sistemas possuem várias vantagens. Tendo em conta que existe uma elevada quantidade de conteúdo, com vários produtos disponíveis *online* muitas vezes os utilizadores têm dificuldades de encontrar o que pretendem, assim como perceber o que efetivamente querem. Desta forma, estes sistemas permitem recomendações úteis e relevantes que adicionam valor à procura do comprador, fazendo-o poupar tempo e fazer escolhas assertivas, de encontro aos seus gostos.

4 Algoritmos de *Machine Learning*

4.1 *K-Nearest Neighbors (kNN)*

O algoritmo do k-vizinho mais próximo é um método de aprendizagem também conhecido por *Lazy Learning*. Este algoritmo aborda o problema de classificar itens novos, comparando-os com a informação que têm guardada em memória utilizando uma função de similaridade (por exemplo: Distância Euclidiana)[12]. Neste processo, determinam-se os *k-nearest neighbors*, ou seja, encontram-se os k-itens mais próximos do novo item a classificar e é-lhe atribuída uma categoria extrapolada a partir das categorias dos seus *nearest neighbors*. Para encontrar os kNN é necessário utilizar uma métrica de semelhança e caso os itens estejam representados por Vector Space Models (VSM), pode ser utilizada a similaridade de cosseno (Cosine similarity (CS)), uma métrica que mede a semelhança entre dois vetores a partir do ângulo que formam entre si [2]. A CS tenta verificar se dois vetores apontam para uma mesma direção. Por outras palavras, um Vector Space Model consiste num conjunto de vetores que contêm informação sobre as características dos itens.

4.2 *Decision Trees*

Este algoritmo tem por base uma estrutura em árvore, baseando-se na decisão por teste, ou seja, qual teste realizar em cada nodo. A complexidade deste algoritmo aumenta de acordo com a quantidade de nodos e do número de ramos da árvore. Este algoritmo de *machine learning* está envolvido nos modelos de aprendizagem, isto é, com um *data set* utilizado como dados de treino, o modelo aprende para em seguida criar uma árvore de decisão para resolver problemas. Este algoritmo é fundamentalmente usado na técnica *Content-based*.

4.3 *Nave Bayes classifier*

Este algoritmo tem por base um modelo probabilístico. Calcula um conjunto de probabilidades, auferindo a frequência e combinação de valores de um *data set*. Baseia-se na probabilidade condicional e no teorema de Bayes, bem como parte do pressuposto que os acontecimentos são independentes [6]. Assim, implica que, um

acontecimento ocorrer ou não, não impede outro de ocorrer. Assumir que os acontecimentos são independentes pode retirar alguma assertividade na recomendação, uma vez que existem acontecimentos que podem efetivamente estar relacionados.

Este algoritmo, para que tenha o seu desempenho aumentado, define dois perfis para cada utilizador: um que é aprendido do histórico do utilizador e outro criado/definido pelo próprio utilizador. Misturar as duas classificações torna-se útil uma vez que, numa fase inicial, existe ausência ou escassez de avaliações históricas. Sendo assim, o perfil aprendido pelo sistema é usado apenas mais tarde na recomendação [1]. É maioritariamente utilizado na técnica de *collaborative filtering* e na *content-based*.

4.4 *Constraint Satisfaction Problem*

Os Constraint Satisfaction Problems (CSPs) são caracterizados por problemas matemáticos definidos com um conjunto de objetos cujo estado dos mesmos deve satisfazer uma série de restrições.

Um problema de satisfação de restrição envolve um conjunto de n variáveis, cada uma representada pelos seus valores de domínio, e um conjunto de restrições. Uma restrição C_i é um subconjunto do produto cartesiano que especifica os valores das variáveis que são compatíveis entre si. Uma solução é atribuir valores a todas as variáveis que satisfaçam todas as restrições e a tarefa é encontrar uma ou todas as soluções. Uma restrição é geralmente representada pelo grupo de todos os conjuntos permitidos por ele.

Os CSPs são bastante utilizados tanto em problemas de pesquisa como em problemas que envolvem Inteligência Artificial. Tal facto, advém da regularidade presente na formulação do mesmo, que de certa forma proporciona uma base comum para análise e resolução de problemas de famílias não relacionadas. Seguem-se alguns exemplos de problemas que utilizam CSPs: o problema das oito rainhas, coloração de grafos, Sudoku, puzzles lógicos, etc.

4.5 *Support Vector Machine*

O objetivo deste algoritmo é encontrar uma plano linear (fronteira de decisão) que separa os dados em duas classes que maximize a margem entre elas. Neste algoritmo os utilizadores e os itens são representados em vetores no espaço e as classificações dadas pelos utilizadores funcionam como variáveis dependentes [3]. Objetivamente, o *support vector machine* é responsável, por exemplo, por separar as classificações positivas das negativas, sendo que o objetivo principal é maximizar a distância entre os pontos mais próximos (vetores que possuem os itens e os utilizadores), de cada uma das classes. Desta forma, na escolha dos itens e utilizador a recomendar poderá existir uma reduzida possibilidade de erro na recomendação, uma vez que a linha que "separa" essas duas classes é "bem definida". Desta forma, percebe-se que este algoritmo é muito utilizados nos sistemas baseados em *content-based recommendation*.

4.6 Artificial Neural Networks

Este algoritmo compreende um conjunto de nodos conectados por ligações, tendo por base a estruturação do cérebro humano, e portanto, a designação dos nodos, neurónios [1]. É composto por três camadas, sendo elas: *input*, *hidden* e *output*. Os dados encontram-se na camada *input*, sendo que o número de neurónios existentes nesta camada é igual ao número de dados que servem de *input*. O processamento dos dados é feito na camada *hidden*, sendo que podem haver várias camadas desta. A *output layer* é a camada na qual o *output* é gerado [7]. É assim utilizado em sistemas *content-based*.

5 Exemplos característicos

O artigo foca-se em sistemas de recomendação no comércio electrónico, na hotelaria e restauração. Existem imensos exemplos da aplicação de sistemas de recomendação no comércio electrónico, desde a Amazon ou eBay. Enquanto na área da hotelaria e restauração existe o Airbnb, o Booking e o TripAdvisor.

A Amazon foi a primeira grande empresa a utilizar um sistema de recomendação, em 1998, sendo pioneira em recomendações para milhões de clientes e tendo um catálogo de milhões de itens. Atualmente, utiliza a filtragem colaborativa *item-to-item*, que permite produzir recomendações de alto nível em tempo real. Este tipo de filtragem combina os itens comprados e classificados pelo utilizador com itens semelhantes. Depois produz uma lista de recomendação para o utilizador com esses itens semelhantes. O algoritmo de recomendação da Amazon cria, de forma eficiente, uma experiência de compra personalizada para o utilizador que permite aumentar o valor médio das encomendas, assim como, da receita gerada por cada cliente [13].

No site do TripAdvisor existe uma área para escrever uma avaliação. Nesta secção, o revisor escreve um texto sobre a sua experiência pessoal num estabelecimento específico, sendo posteriormente solicitado uma seleção de cinco categorias possíveis para classificar o estabelecimento (muito mau, mau, médio, muito bom, excelente). A última classificação permite que o TripAdvisor crie classificações de viajantes para cada atração de hotel, restaurante e outras atrações a visitantes no *site*. Esse processo envolve o cálculo de uma pontuação resumida com base na qualidade, quantidade e idade das avaliações individuais dos viajantes [10]. As técnicas de *Machine learning* são utilizadas para extrair significado dos enormes conjuntos de dados do TripAdvisor. E ainda, é utilizado o processo de filtragem colaborativa que combina interesses específicos dos utilizadores a críticas que incluem conteúdo semelhante - em essência, procurando padrões para posteriormente possibilitar a apresentação de críticas relevantes a cada viajante [8].

6 Conclusão

O sucesso de um negócio depende muito da própria capacidade de aproveitar os sistemas de recomendação a seu favor. Quanto mais relevante for a recomendação,

maiores serão as hipóteses da compra aumentar. Uma vez que estes sistemas são suportados por técnicas, a escolha destas depende da área de negócio na qual se atua, bem como os algoritmos que as suportam.

Quando a utilização de sistemas de recomendação faz sentido, estes constituem vantagens para o promotor, nomeadamente, aumentam a eficiência do controlo do negócio, o envolvimento com clientes e, conseqüentemente, o aumento de vendas e lucros. Para além do mencionado anteriormente, acarreta também várias vantagens para o consumidor. Desta forma, o cliente poupa tempo nas suas escolhas, uma vez que as recomendações vão de encontro àquilo que ele necessita. Assim, os esforços para a implementação de sistemas de recomendação serão compensados com estas mesmas vantagens.

References

1. Ricci, F., Rokach, L.: Artificial Neural Networks Bracha Shapira: Recommender Systems Handbook. 2nd edn. Springer, (2015)
2. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems: An Introduction. 1st edn. Cambridge University Press, (2011)
3. Charu C. Aggarwal: Recommender Systems: The textbook. Springer, (2016)
4. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G. (2010). Hybrid recommendation approaches. In Recommender Systems: An Introduction (pp. 124-142). Cambridge: Cambridge University Press.
5. Isinkaye, F., Folajimi, Y., Ojokoh, B. : Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal(16) .261–273,(2015).
6. Sherekar, S., Patil, T. : Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. International Journal Of Computer Science And Applications (6). 256–261, (2013).
7. Akyuz, I.: Future projection and the sales of industrial wood in Turkey: artificial neural networks. Turkish Journal of Agriculture and Forestry. 368-377, (2019).
8. How TripAdvisor is using AI to "humanize" its review community ,<https://www.phocuswire.com/TripAdvisor-AI-reviews> 16 Oct 2019
9. Bergner, Y., Dröschler, S., Kortemeyer, G., Rayyan, S., Seaton, D., Pritchard, D. E.: Model-Based Collaborative Filtering Analysis of Student Response Data: Machine-Learning Item Response Theory. In: Proceedings of the 5th International Conference on Educational Data Mining, pp. 95–102. Cambridge: MIT, (2012).
10. Jeacle, I., Carter, C.: In TripAdvisor we trust: Rankings, calculative regimes and abstract systems, 293—309. Accounting, Organizations and Society, (2011).
11. Sharma, R., Singh, R.: Evolution of Recommender Systems from Ancient Times to Modern Era: A Survey. Indian Journal of Science and Technology, Vol 9(20),(2016).
12. Notley, S., Magdon-Ismail, M.: Examining the Use of Neural Networks for Feature Extraction: A Comparative Analysis using Deep Learning, Support Vector Machines, and K-Nearest Neighbor Classifiers. In: Computer Science Statistics. <http://arxiv.org/abs/1805.02294>. (2018).
13. The Amazon Recommendations Secret to Selling More Online, <http://rejoiner.com/resources/amazon-recommendations-secret-selling-online/>. Last accessed 15 Oct 2019