

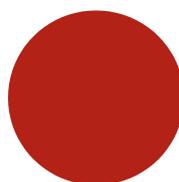
MESTRADO EM CIÊNCIA DA INFORMAÇÃO

Sistema de Recomendação para Percurso Culturalis

Francisca Reis Rodrigues

M
2019

UNIDADES ORGÂNICAS ENVOLVIDAS
FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



Francisca Reis Rodrigues

Sistemas de Recomendação para percursos culturais

Dissertação realizada no âmbito do Mestrado em Ciência da Informação,
orientada pelo Professor Doutor João Pedro Carvalho Leal Mendes Moreira.

Faculdade de Engenharia e Faculdade de Letras

Universidade do Porto

Julho de 2019

Sistema de Recomendação para Percursos Culturais

Francisca Reis Rodrigues

Dissertação realizada no âmbito do Mestrado em Ciência da Informação,
orientada pelo Professor Doutor João Pedro Carvalho Leal Mendes Moreira.

Membros do Júri

Presidente: Professor Doutor Gabriel de Sousa Torcato David

Faculdade de Engenharia - Universidade do Porto

Orientador: Professor Doutor João Pedro Carvalho Leal Mendes Moreira

Faculdade de Engenharia – Universidade do Porto

Arguente: Professor Doutor João Nuno Vinagre Marques da Silva

Faculdade de Ciências - Universidade do Porto

Agradecimentos

Foram várias as pessoas que contribuíram para a realização deste trabalho.

Em primeiro lugar, quero agradecer ao meu orientador, Professor Doutor João Pedro Moreira, pela compreensão das minhas dificuldades, ajuda e orientação durante todo o projeto garantindo um apoio vital para que tudo desse certo.

Quero também agradecer à Professora Doutora Manuela Pinto e ao investigador Rodolfo Matos, pelo auxílio, apresentando-me os projetos da MDUP #IWASHERE e #GPSEngenharia, que foram ponto de partida para esta dissertação, tanto como a prestação do auxílio necessário no início do seu desenvolvimento.

Tenho também de reconhecer os meus colegas de Mestrado que me acompanharam nesta fase e que me encorajaram.

Gratifico os meus amigos que sempre me tentaram entender e ajudar e, sobretudo, ao Carlos pela paciência e apoio que me transmitiu ao longo deste projeto.

Esta realização não seria possível sem a minha família, os meus pais e irmão, e a eles quero agradecer por todo carinho, entendimento e auxílio. Sem eles nada disto seria realizado.

A todos vocês: Obrigada!

Resumo

Sendo uma instituição de memória, com uma vastidão de informação, o museu da Universidade do Porto é um local de importância elevada que deve ser cuidado. A transmissão, partilha e difusão dessa mesma informação para todo o público ao qual se oferece deve ser facilitada.

A aplicação #IWASHERE criada pelo museu pretende ir ao encontro desta difusão da informação, aplicação que conta com a identificação de pontos de interesse e roteiros digitais referentes à cidade do Porto. Com o público-alvo muito vasto e com vantagens denotáveis, é imprescindível a criação de um Sistema de Recomendação que facilite a procura e pesquisa do utilizador.

O objetivo centra-se na organização de todas as ferramentas que permitem recomendar um trajeto novo ou pontos de interesse novos a um utilizador no sistema. Para isso, apresentam-se dois cenários diferentes, ambos usando a filtragem colaborativa. O primeiro faz a semelhança entre utilizadores que percorreram trajetos e o segundo faz a semelhança entre utilizadores que visitaram os pontos de interesse. Cada um destes cenários é formulado em pseudocódigo. A definição do problema de otimização permitirá propor trajetos novos usando esses pontos visitados pelos utilizadores que o utilizador ativo ainda não tenha visitado.

Palavras-chave: #IWASHERE, Filtragem Colaborativa, Pontos de Interesse, Sistema de Recomendação, Otimização

Abstract

Being an institution of memory, with a vast information, the museum of the University of Porto is a place of high importance that must be taken care of. The transmission, sharing and dissemination of this information to all the public to which it is available, should be facilitated.

The #IWASHERE application created by the museum intends to meet this diffusion of information. The application relies on the identification of Points of Interest and digital scripts referring to the city of Porto. With the large target audience and with undeniable advantages, this is essential to create a Recommendation System to facilitate user search.

The goal is to organize all the tools that allow a recommendation of a new route or new Points of interest to a user in the system. For this, two different scenarios are presented, both using collaborative filtering. The first makes the similarity between users who have traveled paths and the second makes the similarity between users who visited points of interest. Each of them is formulated in pseudocode. Defining the optimization problem will allow to recommend new routes using the points visited by users that the active user has not yet visited.

Keywords: #IWASHERE, Collaborative Filtering, Points of Interest, Recommendation System, Optimization

Lista de ilustrações

Figura 1.1 - Árvore de objetivos	4
Figura 2.1 – Filtragem Colaborativa (Adaptado de Rolim et al. 2017, figura 2)	9
Figura 2.2 – Filtragem baseada em Conteúdo (Adaptado de Rolim et al. 2017, figura 1)	11
Figura 2.3 – Exemplo de trajetórias semânticas (Ying et al. 2011, figura 9)	25
Figura 2.4 - Roteiro com pontos de interesse e suas respectivas localizações geográficas	28

Lista de tabelas

Tabela 2.1 – Características dos Sistemas de Recomendação	16
Tabela 2.2 – Uso da Filtragem Colaborativa em semelhanças de trajetos e de POIs	32

Lista de equações

Equação 2.1 - Medida Correlação de Pearson (Aggarwal 2015, pág. 607)	22
Equação 2.2 - Medida Cosseno (Li et al. 2008, equação 7).....	22
Equação 2.3 - Medida Simple_USim (Huang e Gartner 2015, pág. 7).....	23
Equação 2.4 - Medida PIP (Ahn 2008, pág. 42)	24
Equação 2.5 - Medida MSTPS pela média de rácios das partes comuns entre utilizadores (Ying et al. 2010, equação 3)	25
Equação 2.6 - Medida MSTPS pela média de peso de cada trajeto (Ying et al. 2010, equação 4).....	25
Equação 2.7 - Função-objetivo	34
Equação 2.8 - Problema de otimização traduzido	35

Lista de abreviaturas e Siglas

Lista de abreviaturas

DCG	<i>Discounted Cumulative Gain</i>
FC	Filtragem Colaborativa
ID	Identificador
IOT	<i>Internet of Things</i>
KNN	<i>k-Nearest Neighbor</i>
LBS	<i>Location Based Services</i>
LCS	<i>Longest Common Subsequence</i>
MAE	<i>Mean Absolute Error</i>
MDUP	Museu Digital da Universidade do Porto
MSTPS	<i>Maximal Semantic Trajectory Pattern Similarity</i>
NMAE	<i>Normalized Mean Absolute Error</i>
PIP	Proximidade-Impacto-Popularidade
POI(s)	Ponto(s) de interesse
RMSE	<i>Root Mean Squared Error</i>
SR	Sistema de Recomendação

Lista de símbolos

a, b, c – Coeficientes

a1 – utilizador 1

a2 – utilizador 2

C_{a1} - Conjunto de itens avaliados pelos utilizadores a1

C_{a2} - Conjunto de itens avaliados pelos utilizadores a2

EA – Média igual de semelhança

i – Ponto de interesse sugerido ao utilizador u

I – Conjunto de Itens

k – itens

N – Trajeto N

P – Total de pontos de interesse visitados

POI – Conjunto de todos os Pontos de Interesse

POI1 – Ponto de Interesse 1

POI2 – Ponto de Interesse 2

POIs_S – Pois visitados pelos utilizadores em S que não constam de P_u

P_Semelhantes – Pontos de Interesse semelhantes

P_u – Pontos de Interesse visitados pelo utilizador ativo

P_ui – Pontos de Interesse visitados pelo utilizador i

Q – Trajeto Q

ra_{1k} - Avaliações ao item k pelo utilizador a1

ra_{2k} - Avaliações ao item k pelo utilizador a2

S - Total de utilizadores semelhantes

T – Total de trajetos realizados

T_semelhantes – Trajetos semelhantes

T_u – Trajetos realizados pelo utilizador ativo

T_ui – Trajetos realizados pelo utilizador i

u – Utilizador ativo

U – Conjunto de utilizadores

u_i – O utilizador i

$USim$ – Semelhança entre utilizadores

#U – Número total de utilizadores

Índice de conteúdo

1. Introdução	1
1.1. Contexto e motivação	1
1.2. Problemas, objetivos e resultados esperados	2
1.3. Abordagem metodológica	4
1.4. Estrutura da dissertação	4
2. Revisão da Literatura	6
2.1. Sistema de Recomendação	6
2.1.1 Definição	7
2.2. Tipos de Sistemas de Recomendação	8
2.2.1 Filtragem Colaborativa	8
2.2.2 Baseados em conteúdo	11
2.2.3 Baseado em conhecimento	12
2.2.4 Híbridas	13
2.2.5 Demográfico	14
2.2.6 Baseado em comunidade	15
2.2.7 Baseado em Sequências	15
2.3. Problemas nos Sistemas de Recomendação	17
2.3.1 Cold-start	17
2.3.2 Esparsidade de dados	18
2.3.3 Mudança constante de preferências	18
2.3.4 Escala	18
2.3.5 Gray Sheep	19
2.3.6 Privacidade	19
2.3.7 Ataques de shilling	19
2.3.8 Limitações da análise de conteúdo	20
2.3.9 Falta de consciência do contexto	20
2.3.10 Especialização excessiva	20
2.4. Medidas de semelhança entre utilizadores	21
2.4.1 kNN	21
2.4.2 Correlação de Pearson	22
2.4.3 Cosseno/Cosseno ajustado	22
2.4.4 Medida simples de semelhança entre utilizadores (Simple_USim)	22
2.4.5 Distância euclidiana	23
2.4.6 Proximidade-Impacto-Popularidade	23
2.4.7 Semelhança Máxima de Padrões de Trajetórias Semânticas	24
2.5. Medidas de Avaliação em Sistemas de Recomendação	26
2.5.1 Avaliação das previsões, que conferem precisão, ou qualidade destas	26
2.5.2 Conjunto de métricas de recomendação que conferem qualidade	26
2.5.3 Qualidade da lista de recomendações	27
2.5.4 Métricas de diversidade	27
2.6. Implementação de um Sistema de Recomendação na aplicação do Museu Digital da Universidade do Porto	28
2.7 – Complicações ao longo do projeto	29
3. Metodologia	30
3.1. Porquê utilizar a Filtragem Colaborativa?	30
3.2. Dois panoramas diferentes	31
3.2.1 Filtragem colaborativa por trajetos	32

3.2.2 Filtragem colaborativa de Pontos de Interesse.....	32
3.2.3 Pseudocódigo.....	33
3.3. Otimização do problema.....	33
3.4. Contribuições do projeto.....	35
4. Conclusões e perspectivas de desenvolvimento	36
4.1. Perspetivas futuras	36
5. Referências Bibliográficas	38
Anexos	41
Anexo 1 - Filtragem Colaborativa por trajetos	41
Anexo 2 - Filtragem colaborativa por pontos de interesse.....	42
Anexo 3 - Argumento de maximização da função	42

1. Introdução

1.1. Contexto e motivação

A Universidade do Porto contém património que está ligado à cidade e ao seu desenvolvimento, pretendendo a sua divulgação social. Esta universidade tem uma formação com grande incidência na investigação e na valorização do conhecimento social e económico e ativa participação no desenvolvimento da comunidade na qual se encontra inserida.

A preocupação do Museu Digital da Universidade do Porto (MDUP) detém-se nesse mesmo ponto, pretendendo o zelo pela gestão apropriada do seu espólio, bem como sua difusão e acessibilidade aos interessados. A dinamização da troca de conhecimento, a gestão das coleções, a disponibilização de visualização no *website* da UP e a inventariação de todo o acervo, são competências que o museu deve ter e disponibilizar. A estratégia do Museu Digital da Universidade do Porto é que este seja um lugar de troca de conhecimento e partilha para dentro e fora da comunidade académica.

Então, tendo como escopo principal a transmissão de informação ao seu público, de forma a satisfazer as necessidades dos interessados, este museu criou uma aplicação e uma base de dados: #IWASHERE e #GPSEngenharia, respetivamente (Pinto et al. 2016).

A aplicação #IWASHERE conta com a identificação de pontos de interesse e roteiros digitais referentes ao património universitário edificado. Esta aplicação, a partir da partilha de experiências de determinados locais, bem como de submissões de vídeos, textos ou fotografias, sugere roteiros pessoais com determinados Pontos de Interesse (POIs) em cada um. Cada roteiro constitui um conjunto de pontos de interesse a si referentes. A plataforma ainda disponibiliza percursos que são gerados a partir de uma lista de POIs que se encontram diretamente ligados. A criação desta aplicação funciona como orientadora na cidade do porto e permite a aprendizagem da história local, incentivando a partilha de conhecimento e experiências acerca de determinados locais de interesse e património. Já a base de dados #GPSEngenharia inclui conteúdos de locais que se relacionam com a Engenharia na região norte do Portugal, como pontes, edifícios, entre outros.

Estes POIs ou percursos culturais de diferentes rotas vão ajudar o visitante a organizar e gerir o seu caminho, segundo o interesse, necessidade e preferência de cada um. Destes percursos podem aproveitar todo o tipo de pessoas, individual ou grupal, como estudantes, famílias, turistas, etc. Estes últimos beneficiam pela oportunidade de conhecer melhor a cidade e pontos turísticos que possam ter interesse dentro dos caminhos sugeridos. A universidade do Porto e seus museus podem representar uma mais-valia na expansão turística do Porto, consoante a qualidade histórica que os museus possuem (Pinto et al. 2016). É assim fundamental referir que o património universitário e o turismo podem representar uma vantagem significativa na divulgação da cidade do Porto.

As rotas que contém um ou vários POIs são feitas por cada utilizador, e os dados recolhidos correspondem, essencialmente, à sucessão de POIs, criando um padrão de sequências. Essas sequências vão permitir, através da filtragem colaborativa, a indicação de sequências ou subsequências a outros utilizadores novos no sistema.

Torna-se crucial haver uma solução para o excesso de informação, restringindo aquilo que é necessário para o utilizador, obtendo um carácter eficiente. Esta solução será possível com a implementação de um Sistema de Recomendação (SR). O Sistema de Recomendação é um conjunto de técnicas que ajudam o utilizador na filtragem dessa mesma informação e é o estudo das preferências e necessidades por parte de cada utilizador ou grupo destes.

Como não houve utilizadores suficientes, não existiram dados referentes às suas atividades. Esta escassez dificultou a realização deste projeto, forçando a ser utilizada uma nova abordagem.

Para isto, são apresentados dois cenários diferentes – filtragem colaborativa por trajetos e filtragem colaborativa por POIs – que permitirão encontrar o top S utilizadores de acordo com a semelhança entre trajetos ou POIs, de modo a propor trajetos novos usando POIs visitados pelos utilizadores semelhantes que o utilizador ativo ainda não tenha visitado. Cada uma destas abordagens é, seguidamente, formulada em pseudocódigo.

A realização desta dissertação é uma alavanca para a criação destas recomendações, onde estes dois cenários apresentados descrevem medidas que assemelham utilizadores, sugerindo as mais relevantes para utilizar aquando a criação do Sistema de Recomendação. Estas medidas são de filtragem colaborativa baseada em memória, que assemelham utilizadores. A otimização do problema permitirá potencializar o que deve ser recomendado.

Durante este estudo, será mencionado o termo “utilizador ativo” usualmente, que representa o utilizador alvo das recomendações. Caso o utilizador não tenha feito nenhum percurso, estamos perante o problema *Cold-start* (ver 2.3.1), em que não existem dados sobre aquilo que visitou, impossibilitando saber o seu histórico, não sendo possível fazer a semelhança com outros visitantes.

É também importante referir que a noção de item em Sistemas de Recomendação é, para o contexto desta dissertação, a mesma coisa que POIs.

Em suma, nesta presente dissertação denominada “Sistemas de Recomendação para percursos culturais”, realizada no âmbito do Mestrado em Ciência da Informação na Faculdade de Engenharia, pretende-se fazer recomendações mais úteis e precisas aos utilizadores, de modo a satisfazer as suas necessidades iniciais.

1.2. Problemas, objetivos e resultados esperados

O objetivo principal deste projeto é dar ferramentas para criação de um Sistema de Recomendação que permita recomendar percursos/rotas, definido e adaptado para cada utilizador, segundo o seu perfil e as suas características.

Deste modo, é necessário cumprir os seguintes passos:

- Identificar dados disponíveis. Estes dados são relacionados com as preferências e necessidades do utilizador, a partir das sequências de POIs que realizaram;
- Identificar medidas de semelhança entre utilizadores;
- Usar essas medidas para comparar trajetos realizados e POIs visitados pelos utilizadores;
- Descobrir quais utilizadores com trajetos e com POIs semelhantes;
- Selecionar os top S utilizadores semelhantes ao utilizador ativo;
- Otimizar o problema;
- Propor trajetos novos usando POIs visitados pelos utilizadores semelhantes que o utilizador ativo ainda não tenha visitado.

A partir destes objetivos é esperado que:

- O conjunto de dados pesquisado e utilizado forneça informação suficiente para a realização deste trabalho;
- Haja uma partilha/transmissão de informação maior em diferentes comunidades, em meio digital.

Na figura seguinte é apresentada a árvore de objetivos:

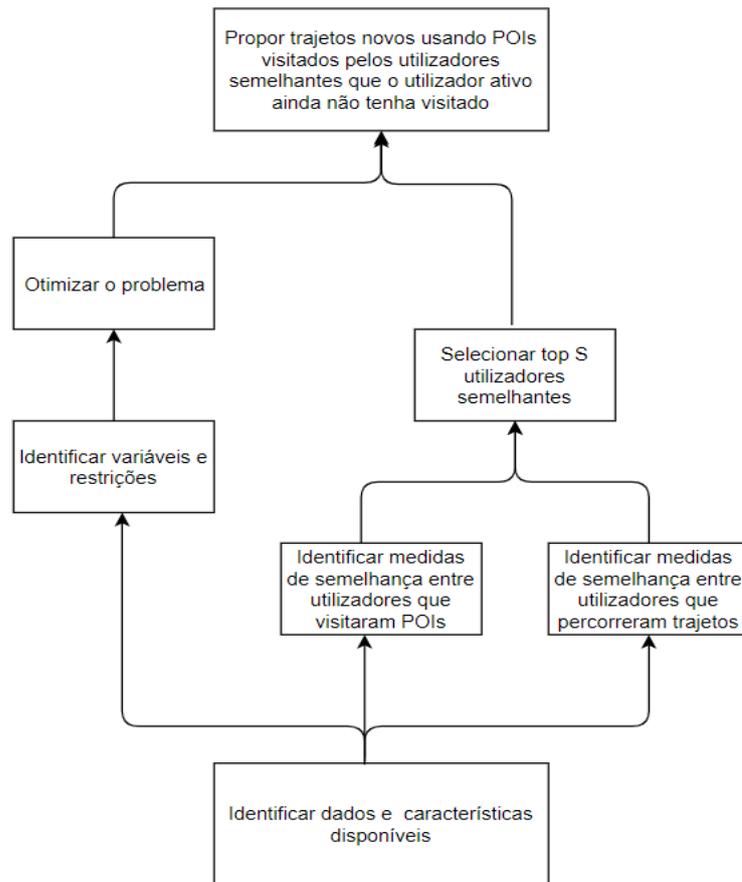


Figura 1.1 - Árvore de objetivos

1.3. Abordagem metodológica

É um projeto maioritariamente exploratório, em que é formulado o problema e são definidas as hipóteses com precisão, bem como técnicas adequadas. Funcionando como base para outras finalidades, esta dissertação não segue nenhuma metodologia fechada.

Para fundamentar o presente trabalho foram pesquisadas fontes secundárias como artigos com base científica. Foram também selecionados alguns livros não lidos na íntegra, mas com teor importante no decorrer da escrita da dissertação.

Os dados recolhidos da aplicação #IWASHERE e do projeto #GPSEngenharia contêm informação sobre os POIs como nome, texto resumo, rua, coordenadas GPS e observações.

1.4. Estrutura da dissertação

Esta dissertação inicia-se com a contextualização do tema e a motivação que levou ao estudo e reflexão deste. São apresentados os problemas tanto a nível metodológico como teórico tendo em foco mostrar o problema principal de estudo. A abordagem metodológica expõe qual o plano

metodológico utilizado e por fim, são descritos os passos de estrutura da escrita da dissertação.

No segundo capítulo é apresentada a Revisão da Literatura descrevendo, em termos teóricos, todos os pontos importantes de estudo. Deste modo, é descrito o que é um Sistema de Recomendação, os seus tipos e problemas, mostrando soluções plausíveis para esses problemas. Mostrou-se importante identificar as medidas de semelhança entre itens e as medidas de avaliação em Sistemas de Recomendação.

No capítulo 3 descreve-se a metodologia, expondo os resultados obtidos.

Por último são apresentadas as conclusões retiradas com a realização deste trabalho e possíveis projetos futuros.

A dissertação finaliza com a apresentação das referências bibliográficas.

2. Revisão da Literatura

“Museu é uma instituição permanente sem fins lucrativos, ao serviço da sociedade e do seu desenvolvimento, aberta ao público, que adquire, conserva, investiga, comunica e expõe o património material e imaterial da humanidade e do seu meio envolvente com fins de educação, estudo e deleite” (ICOM 2015).

É uma organização sem fins lucrativos que coleciona, analisa, preserva e apresenta objetos pertencentes ao património natural e cultural, com o intuito de aumentar a quantidade e qualidade do conhecimento. Museu é entendido como uma instituição prestadora de serviços da sociedade, o qual possui os elementos que lhe permitem formar a consciência das comunidades a que ele serve (Poulot 2013).

Segundo Solima, Peruta e Maggioni (2016) os museus atentam na sustentação e diversidade constante de consumo cultural, num espaço fechado, finito e definido. Visto o grande número de objetos que representam a sua coleção, a preocupação fulcral será determinar qual conhecimento deve ser transmitido.

O comportamento dos visitantes demonstra um papel importante durante a sua visita. Este comportamento é determinado por três variáveis: as características pessoais dos visitantes, os fatores de tipo estrutural e uma combinação dos dois. Estas características pessoais referem-se ao nível de escolaridade, interesses específicos, tempo disponível, aspetos sociais, idade, sexo e nacionalidade, bem como em aspetos da coleção como importância dos objetos, período, técnicas, temas, etc. Estas três variáveis podem afetar os processos de tomada de decisão nos envolvidos na visita.

Ao utilizar as tecnologias de *Internet of Things* (IOT) (Internet das coisas) em SR, os museus podem usar os seus recursos de forma eficiente, gerindo as relações dos visitantes com os objetos e com o sistema de conhecimento pertencente aos funcionários do museu.

Atualmente, os SR têm um papel importante na tomada de decisão sobre a escolha daquilo que o utilizador quer fazer, consoante o itinerário escolhido, o tempo que quer gastar ou mesmo na sequência de lugares que pretende visitar (Cai, Lee e Lee 2018).

2.1. Sistema de Recomendação

De modo a compreender um Sistema de Recomendação, é essencial conhecer os seus conceitos associados, os quais se designam: item, utilizador, avaliação e transação.

O termo item é geralmente usado para indicar o que o sistema recomenda aos utilizadores, ou seja, é o objeto que é recomendado. Pode ser caracterizado segundo a sua complexidade ou valor de utilidade. É positivo se for útil para o utilizador e negativo se não for apropriado para o mesmo (Recommender Systems Handbook 2015). Os itens podem ser, por exemplo, filmes, notícias, livros, vídeos, etc., ou seja tudo o que poderá ser recomendado a um utilizador.

Para Schafer, Frankowski e Sen (2014), o utilizador define-se como qualquer indivíduo que usufrua do sistema com o intuito de receber informação (isto é, recomendações) e dar avaliações.

Os mesmos autores descrevem avaliação como uma associação entre duas coisas: utilizador e item, normalmente significando algum valor. É, então, a contagem entre o utilizador e o item. As avaliações podem ser de diferentes tipos: avaliação em escala, ou seja, de 1 a 5, em que 1 é muito mau e o 5 é excelente; avaliação do tipo binário, em que resposta será “concordo” ou “discordo”.

A transação é a interação entre o utilizador e o sistema (Recommender Systems Handbook 2015).

2.1.1 Definição

O estudo sobre Sistemas de Recomendação, segundo Adomavicius e Tuzhilin (2005) começou a ser um tema emergente em meados de 1990. Os Sistemas de Recomendação representam as preferências e necessidades dos utilizadores com o propósito de sugerir itens para comprar ou examinar, bem como descobrir itens novos (Burke 2002).

Frequentemente, estes sistemas regem-se pelas avaliações e opiniões dadas pelos utilizadores no sistema, sendo estes próprios que sugerem possíveis caminhos, pois as recomendações são feitas a partir dos perfis dos utilizadores (Batet et al. 2012). Através do que o utilizador faz dentro do sistema, estes caminhos – pontos de interesse – são recuperados, analisados e filtrados. Os POIs são listas de pontos de interesse que vão ao encontro das necessidades dos utilizadores.

Em geral, os SR são, então, definidos como sistemas de suporte que ajudam os utilizadores a encontrar informações, produtos e serviços, agregando e analisando sugestões de outros utilizadores como avaliações e atributos do utilizador.

Segundo Bobadilla et al. (2013), o processo para gerá-los é baseado na combinação entre os seguintes pontos:

- O tipo de dados disponível na base de dados, como por exemplos, as avaliações, a informação registada do utilizador, características e conteúdo dos itens que podem ser avaliados, etc;
- O algoritmo de filtragem utilizado;
- O modelo escolhido (baseado em memória ou baseado em modelo);
- As técnicas implementadas: Redes *Bayesianas*, modelos *fuzzy*, *Singular Value Decomposition*, abordagens probabilísticas, etc;
- O nível de dispersão e escala na base de dados;
- Tempo e memória consumidos no sistema;
- Qualidade dos resultados;

- O objetivo previsto.

Os SR têm um conjunto de vantagens associadas (Recommender Systems Handbook 2015) :

- Melhora a compreensão sobre o que os utilizadores precisam;
- Aumenta o número de itens vendidos;
- Aumenta a diversificação de itens vendidos;
- Aumenta a fidelidade do utilizador e conseqüentemente a sua satisfação.

Anandhan et al. (2018) acrescentam que os SR:

- Podem reduzir o tempo e custo gasto pelos utilizadores e implementar estratégia de processo, qualidade e tomada de decisão pelos fornecedores.

Os SR manuseiam o excesso de informação (Khusro, Ali e Ullah 2016) e são desenvolvidos para reduzir o problema da quantidade enorme de informação produzida na internet (Bobadilla et al. 2011).

Resumindo, ajuda os utilizadores na pesquisa do seu caminho nas grandes bases de dados e catálogos, filtrando e sugerindo itens relevantes nas contas das preferências dos utilizadores (gostos, interesses, prioridade, etc) (Bellogín e Sánchez 2017). O seu objetivo principal é, então, dar recomendações precisas aos utilizadores (Madadipouya e Chelliah 2017).

2.2. Tipos de Sistemas de Recomendação

As quatro diferentes abordagens usadas no desenvolvimento dos Sistemas de Recomendação são: filtragem colaborativa, filtragem baseada em conteúdo, filtragem híbrida e filtragem baseada em conhecimento (Anandhan et al. 2018). Aggarwal (2016) acrescenta as filtragens baseada em utilidade e em comunidade.

2.2.1 Filtragem Colaborativa

A Filtragem Colaborativa (FC), segundo Aggarwal (2016), é um levantamento das preferências dos utilizadores na forma de comportamentos de compra e avaliações de forma colaborativa para todos eles beneficiarem. Burke (2002) considera ser a tecnologia mais madura e afirma que é a técnica mais popular e largamente usada. Após haver uma série de avaliações, é necessário encontrar utilizadores semelhantes, selecionar esses utilizadores segundo a sua proximidade de gostos, de modo a criar um vizinho e selecionar o grupo de itens a ser recomendados (Batet et al. 2012).

É chamada “correlação pessoa-pessoa” e afirmam que se calcula semelhanças na história de avaliações dos utilizadores (Recommender Systems Handbook 2015).

Este tipo de sistema pode usar semelhanças entre itens e entre utilizadores, dependendo, essencialmente do número existente de cada um. Se o número de itens for inferiores ao número de utilizadores, então a semelhança entre itens deve ser usada, de modo a não se fazer muitos

cálculos.

O objetivo principal é fazer recomendações aos utilizadores com base nos comportamentos das compras de utilizadores com gostos semelhantes (Aggarwal 2016). Para isso, cálculo de padrões é útil, onde a ideia é agrupar os dados em segmentos e determinar os padrões nesses segmentos (Aggarwal 2016). Schafer, Frankowski e Sen (2014) assumem que pessoas com os mesmos gostos vão dar avaliações a coisas semelhantes.

A FC agrega avaliações e recomendações dos objetos, reconhecendo o que há em comum dos utilizadores e gerando novas recomendações a partir da comparação das avaliações dos utilizadores (Burke 2002).

A imagem 2.1 esclarece como é feita a filtragem colaborativa.

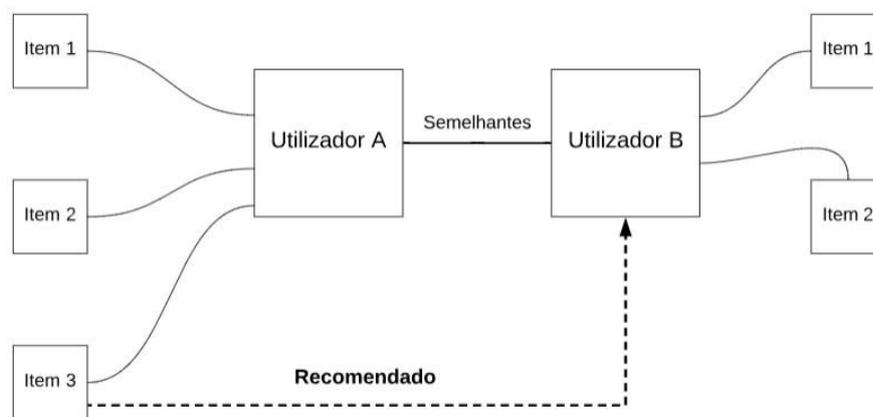


Figura 2.1 – Filtragem Colaborativa (Adaptado de Rolim et al. 2017, figura 2)

As técnicas de filtragem colaborativa têm duas categorias: baseadas em modelo e baseadas em memória (também chamadas de vizinhança ou heurística) (Bellogín e Sánchez 2017; Recommender Systems Handbook 2015). A primeira constrói padrões das diferentes iterações utilizador/item para gerar previsões automáticas. A segunda baseia-se na semelhança entre utilizadores nas suas escolhas dos itens, ou na semelhança dos próprios itens.

A filtragem colaborativa baseada em memória significa fazer previsões baseadas na coleção inteira das interações, com semelhanças entre utilizadores e itens (Adomavicius e Tuzhilin 2005).

Outra designação dada por Hernando et al. (2016) e por Bellogín e Sánchez (2017) é “o vizinho mais próximo”, utilizando o algoritmo associado a esta designação: *k-nearest neighbor* (kNN). Este outro nome é dado porque este tipo de filtragem explora as semelhanças para classificar os utilizadores e itens e usa o mais próximo para gerar recomendações. O algoritmo é usado para prever os gostos dos utilizadores por significados de medidas de semelhança diferentes.

Anandhan et al. (2018), Recommender Systems Handbook (2015) e Schafer, Frankowski e Sen (2014) apresentam um conjunto de vantagens associadas à filtragem colaborativa, sendo as

seguintes:

- Dar uma lista de recomendações baseadas nas interações prévias dos utilizadores e entre estes e os itens;
- Explorar um conjunto de utilizadores que têm história de concordância com o utilizador alvo, consoante a partir da avaliação do conjunto de itens semelhantes entre esses utilizadores;
- Podem implementar precisão na recomendação, quando calculada a semelhança entre os interesses dos vizinhos de cada utilizador;
- Ajudar na pesquisa de novos itens que possa gostar;
- Há um aviso de um item particular;
- Ajudar a encontrar um utilizador (ou alguns) que possa gostar;
- Ajudar o grupo a encontrar algo novo que possa gostar;
- Ajudar a encontrar uma mistura entre itens novos e antigos;
- Ajudar com as tarefas que são específicas neste domínio.

Os autores Binucci et al. (2017) e Anandhan et al. (2018) apresentam um conjunto de desvantagens:

- Problema da dispersão de dados, de escala e *cold-start*;
- Limitado nas preferências dos utilizadores a partir das avaliações de cada novo utilizador;
- Limitado na recomendação do novo item que é apenas confiado em preferências dos utilizadores;
- Dificuldades na aprendizagem.

Baseado em memória

Adomavicius e Tuzhilin (2005) referem-se à memória como heurística. A filtragem colaborativa baseada em memória significa fazer previsões baseadas na coleção inteira das interações, com semelhanças entre utilizadores e itens.

Outra designação dada por Hernando et al. (2016) e por Bellogín e Sánchez (2017) é “o vizinho mais próximo”, utilizando o algoritmo associado a esta designação: *k-nearest neighbor* (kNN). Este outro nome é dado porque este tipo de filtragem explora as semelhanças para classificar os utilizadores e itens e usa o mais próximo para gerar recomendações. O algoritmo é usado para prever os gostos dos utilizadores por significados de medidas de semelhança diferentes.

Os algoritmos baseados em memória podem ser distinguidos em: baseados nos utilizadores e baseados nos itens. Esta distinção pressupõe várias medidas de semelhança. Os algoritmos baseados nos utilizadores procedem vendo o S de utilizadores mais semelhantes com o utilizador

ativo (vizinhança). Esta semelhança deve identificar a distância ou correlação entre dois utilizadores.

Já os algoritmos baseados em itens procedem com os S itens que são semelhantes aos itens vistos e avaliados. Segundo Sarwar o objetivo principal é utilizar uma matrix utilizador-item, a qual permite prever a avaliação de um utilizador a um item não visto anteriormente. Assim, um utilizador estará interessado em ver itens semelhantes aos que viu anteriormente evitando aqueles que não gostou no passado. Primeiro, o algoritmo calcula S itens mais semelhantes e identifica o conjunto de itens candidatos a serem recomendados, removendo também os itens que o utilizador já comprou.

2.2.2 Baseados em conteúdo

Sistemas de recomendação baseados em conteúdo tentam combinar itens que são semelhantes com o que o utilizador gostou no passado, baseando-se no histórico de itens avaliados pelo utilizador (Aggarwal 2016; Recommender Systems Handbook 2015). A semelhança é feita pela análise de conteúdo dos próprios itens (Recommender Systems Handbook 2015). O objetivo principal é identificar as características comuns dos itens que receberam uma avaliação favorável por um utilizador e recomendar a esse mesmo utilizador itens com características comuns/partilhadas com o que ele avaliou anteriormente. A filtragem baseada em conteúdo consegue prever relevância mesmo para os itens sem avaliação, precisa apenas do conteúdo para analisar (Schafer, Frankowski e Sen 2014).

Deste modo, segundo Khusro, Ali e Ullah (2016), é necessário seguir os dois seguintes passos:

- 1º Criar um perfil do utilizador usando o seu feedback e avaliações sobre outros itens;
- 2º O perfil do utilizador é comparado com outras características de itens e os itens que fazem correspondência são recomendados.

A figura que se segue apresenta como é feita a filtragem baseada em conteúdo.

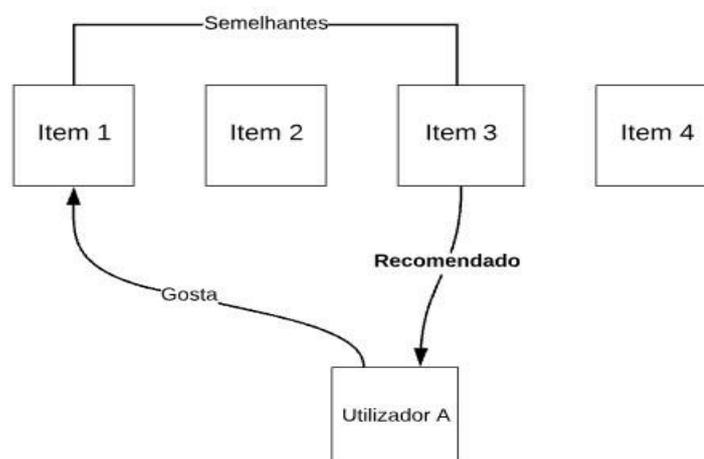


Figura 2.2 – Filtragem baseada em Conteúdo (Adaptado de Rolim et al. 2017, figura 1)

Anandhan et al. (2018) apresentam um conjunto de benefícios e desvantagens associadas à filtragem baseada em conteúdo.

Este tipo de filtragem tem como vantagens:

- Explorar a informação textual que é dada por cada utilizador como comentário/*feedback* em qualquer domínio;
- Usar diferentes métodos para representar os interesses dos utilizadores em formas de pequeno texto;
- Realizar bom trabalho das técnicas de recuperação da informação na extração de características de cada documento de texto.

No entanto:

- A semelhança do texto é um problema no áudio e vídeo relacionados com o domínio de grande conteúdo;
- Há dificuldade em dar atributos para limitar os recursos;
- O conjunto dos itens recomendados pode ser óbvio e homogéneo, problema que leva à excessiva especialização.

2.2.3 Baseado em conhecimento

Estes métodos destacam-se em itens que não são comprados com muita frequência pelos utilizadores. São baseados nas características dos itens que conhecem as necessidades e preferências dos utilizadores e como esse item é útil (Recommender Systems Handbook 2015). Têm a ver com as exigências dos utilizadores e incorporam apenas um número limitado de dados do histórico (Aggarwal 2016). No entanto, o mesmo autor refere que é difícil capturar os interesses do utilizador apenas observando o histórico.

Alguns sistemas baseados em conhecimento podem usar atributos demográficos do utilizador e dos itens, podendo até codificar relações entre os atributos de ambos (Aggarwal 2016).

Alyari e Navimipour (2018) afirmam que usar o modelo baseado em conhecimento é um processo interativo forte, no entanto é difícil de caracterizar e não é possível obter a iniciativa de novos clientes.

Baseado em utilidade

Este tipo de filtragem pode ser visto como um caso específico de SR baseados em conhecimento (Aggarwal 2016). Sistemas baseados em utilidade têm como base os perfis de preferências dos utilizadores (Recommender Systems Handbook 2015). Aggarwal (2016) define utilidade nas características do produto de modo a calcular a probabilidade do utilizador gostar de um item. Segundo este mesmo autor, o desafio principal deste sistema é definir uma função de utilidade apropriada para o utilizador em questão.

2.2.4 Híbridas

Sistemas de recomendação híbridos combinam o poder de todos os diferentes tipos de recomendações (Aggarwal 2016), com o objetivo de melhorar o desempenho (Burke 2002). Adomavicius e Tuzhilin (2005) dizem que estes métodos combinam métodos colaborativos e baseados em conteúdo, ajudando a diminuir/proibir certas limitações destes métodos.

Segundo os mesmos autores, as diferentes maneiras de combinar os dois métodos para gerar híbrido são:

- Implementar métodos separadamente e combinar as suas previsões;
- Incorporar algumas características baseadas em conteúdo na abordagem colaborativa;
- Incorporar algumas características colaborativas na abordagem baseada em conteúdo;
- Construir um modelo geral que incorpora características de ambos os métodos.

Os métodos utilizados para esta técnica são os seguintes (Anandhan et al. 2018; Burke 2002):

- Ponderada: as pontuações das várias técnicas de recomendação são combinadas para produzir uma única recomendação;
- Alterada: o sistema troca entre as técnicas de recomendação dependendo da situação atual;
- Mistura: as recomendações das diferentes técnicas são apresentadas ao mesmo tempo;
- Combinação de características: todos os dados são postos juntos num simples algoritmo de recomendação;
- Cascata: uma recomendação refine todas as recomendações dadas por outrem;
- Características de argumentação – o *output* de uma técnica é usado como *input* de outra;
- Meta-nível – o modelo aprendido por uma recomendação é usado como *input* de outra.

A figura 2.3 apresenta o método híbrido, numa junção entre as três filtragens anteriormente descritas.

Método Híbrido	
Filtragem Colaborativa	Filtragem Baseada em Conteúdo
Independência dos utilizadores	Recomendação de itens dispersos e inesperados
Histórico do consumo	Histórico de avaliações em comum
Indicação dos itens ainda não avaliados por qualquer utilizador	Descoberta de novos relacionamentos entre utilizadores

Figura 2.3 - Filtragem baseada em Método Híbrido (Figura do autor 2019)

O método híbrido possui as seguintes vantagens (Alyari e Navimipour 2018):

- Superar as limitações de outros tipos de recomendação;
- Superar problemas como a dispersão;
- Proporcionar flexibilidade.

E as seguintes desvantagens:

- Grande complexidade e despesa na implementação;
- Incapacidade de dar recomendações agradavelmente surpreendentes.

2.2.5 Demográfico

Este método define-se pelas classes demográficas, como por exemplo: idade, nível de escolaridade, sexo, entre outros (Batet et al. 2012; Burke 2002). Acrescentam que recomendações diferentes devem ser geradas para nichos de diferentes demografias.

As técnicas demográficas são semelhantes às técnicas de filtragem colaborativa nas suas abordagens, mas diferentes nas suas características, isto é, na primeira importa os dados demográficos e na segunda as avaliações, mas ambas comparam utilizadores (Burke 2002).

Segundo Alyari e Navimipour (2018), a filtragem demográfica é usada para identificar os tipos de utilizadores que gostam de determinado objeto, no entanto é baseada na generalização do interesse dos utilizadores.

2.2.6 Baseado em comunidade

Estes tipos de sistemas têm como objetivo recomendar itens baseados nas preferências dos amigos do utilizador. Isto é, a recomendação é feita pelas avaliações que são dadas pelos amigos dos utilizadores. As pessoas têm mais tendência a acreditar nas recomendações dos seus amigos do que de um indivíduo, apesar de singular, anónimo (Recommender Systems Handbook 2015).

2.2.7 Baseado em Sequências

Os sistemas de recomendação em sequência são nem mais nem menos do que a recomendação lógica de uma série de itens. É representado por uma sequência de pontos $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$, em que cada p pode ser, por exemplo, museu, jardim, biblioteca, entre outros. Este tipo de Sistema de Recomendação integra várias características como geografia, cultura, lazer, etc..

Os algoritmos de frequências, essencialmente, demonstram encontrar todas as subsequências frequentes, isto é, as subsequências que ocorrem com frequência num conjunto de sequências.

O *PrefixSpan* é um algoritmo útil para determinar sequências mais frequentes, isto é, num caminho, o algoritmo calcula a frequência de visitas de um determinado POI e vai recomendar o POI mais frequentemente visitado (Aggarwal 2015). Dado um conjunto de POIs a, b, c, d , *PrefixSpan* vai calcular a frequência de vezes que cada um desses POIs foi visitado e irá recomendar o que tiver maior número de frequência.

O algoritmo *Generalized Sequential Pattern* (GSP) também pretende encontrar todos os itens frequentes que formam o conjunto de um único item de sequências frequentes. Inicialmente, o algoritmo, então, encontra todos os itens frequentes e depois apaga todos os que não são frequentes. O algoritmo *Apriori* tem o mesmo objetivo.

Existe também o algoritmo *Location-Item-Time PrefixSpan* (*LIT-PrefixSpan*), descrito por Tsai e Lai (2014), onde sugere um percurso de acordo com requisitos e limitações dos visitantes (gostos, tempo disponível, entre outras).

Já Bellogín e Sánchez (2017) explicam o algoritmo *Longest Common Subsequence* (LCS) usado para determinar a sequência de caracteres, técnica que compara a semelhança das sequências consoante o comportamento dos utilizadores (ver secções 2.4.7 e 3.2.1). Aqui a avaliação é apenas feita segundo se o utilizador visitou ou não o POI.

Existem poucos estudos sobre recomendações de percursos completos.

A tabela 2.1 apresenta um resumo dos tipos de Sistemas de Recomendação apresentados, segundo uma perspetiva de Burke (2002). Apesar da informação da tabela ser antiga, mesmo podendo haver outras categorizações mais atuais, são, eventualmente, menos consensuais.

Tabela 2.1 – Características dos Sistemas de Recomendação

Sistema de Recomendação	Base	Input	Processo
Filtragem Colaborativa	Avaliações de U para itens em I .	Avaliações de u para itens em I .	Identificar utilizadores em U que são semelhantes a u , e extrapolar pelas suas avaliações o item i .
Baseado em Conteúdo	Características dos itens em I .	Avaliações de u para itens em I .	Classificar o comportamento de avaliação de u e utilizar em i .
Demográfico	Informações demográficas sobre U e as suas avaliações a I .	Informações demográficas de u .	Identificar utilizadores demograficamente semelhantes a u e extrapolar pelas suas avaliações o item i .
Baseada em Conhecimento	Características dos itens em I . Conhecimento de como esses itens se adequam às necessidades do utilizador.	Descrição das necessidades ou interesses de u .	Inferir a correspondência entre i e as necessidades de u .

Adptado de Burke, 2002, tabela 1

2.3. Problemas nos Sistemas de Recomendação

2.3.1 Cold-start

Um dos problemas relativo aos Sistemas de Recomendação é o *cold-start*, contratepo sobre o qual vários autores se debruçam. Segundo Hernando et al. (2016), o *cold-start* divide-se em duas variantes: novo item e novo utilizador. Ou seja, quando um novo utilizador se regista no sistema, dificulta a criação de novas recomendações visto que este ainda não fez nenhuma avaliação. Na carência de avaliações de um utilizador, é difícil dispor de recomendações precisas (Adomavicius e Tuzhilin 2005). Os mesmos autores acreditam que novos itens estão sendo constantemente adicionados aos sistemas e quando isto acontece, ocorre outra adversidade, visto que como este novo item ainda não foi avaliado, é difícil que o sistema recomende este mesmo item a qualquer utilizador (Hernando et al. 2017). Para adaptar-se ao utilizador, o sistema necessita do passado deste com informações sobre aquilo que ele gostou (Recommender Systems Handbook 2015). A estas duas variantes descritas, Bobadilla et al. (2013) e Madadipouya e Chelliah (2017) acrescentam o problema do novo sistema que acontece quando este está a ser construído e não há dados suficientes.

Schafer, Frankowski e Sen (2014) e Madadipouya e Chelliah (2017) assumem que o *cold-start* pode ser resolvido das seguintes maneiras:

- Pedir ao utilizador para, no início, avaliar alguns itens diversificados, isto é, devem especificar explicitamente as suas preferências ou avaliar alguns itens antes de estar aptos para entrar no sistema;
- Pedir ao utilizador que mostre claramente os seus gostos;
- Sugerir itens baseados na demografia do utilizador (sexo, idade, grau de escolaridade, entre outros).

Hernando et al. (2016) referem algumas formas de resolução do problema do novo utilizador, de modo a que este dê algumas das suas informações à partida para o sistema conseguir fazer recomendações precisas. Deste modo, estas formas que oferecem grande informação sobre o utilizador serão:

- Registo do utilizador através das suas redes sociais;
- Desenho de uma questão de partida sobre alguns itens que o utilizador gosta;
- É necessário haver pessoas do staff a avaliarem cada item novo no sistema, para solucionar o problema do novo item.

Em 2017, Binucci et al. voltam a abordar a demografia do utilizador, recomendando itens segundo as suas classes, sendo esta uma variante da filtragem colaborativa.

Vários autores dedicam-se no sentido de procurar soluções para este problema pois a filtragem colaborativa tem características ótimas, no entanto com estes defeitos que devem ser

ultrapassados. Desta forma, Guo, Jie e Thalmann (2014) apresentam um novo método chamado “*merge*” que incorpora informação de confiança no fornecimento de recomendações. As classificações dos vizinhos confiáveis de um utilizador são misturadas para complementar e representar as preferências do utilizador e para encontrar novos utilizadores com preferências semelhantes (por exemplo, utilizadores semelhantes). Çano e Morisio (2017) apresentam combinar FC com técnicas como *fuzzy logic*. Este autor também explica que para resolver o *cold-start* deve-se usar cálculo de regras no item ou usar dados para encontrar relações que possam compensar a falta de avaliações. Para resolver o problema da esparsidade deve-se usar as poucas avaliações existentes ou certas características do item para gerar extra *pseudo* avaliações, com *matrix factorization* e *dimensionality reduction*.

2.3.2 Esparsidade de dados

A dispersão dos dados detém-se com a falta de quantidade de avaliações que é feita a um determinado item, dificultando a recomendação desse mesmo item. Ou seja, segundo Adomavicius e Tuzhilin (2005) geralmente, há poucas avaliações dos itens feitas pelos utilizadores. O sucesso da recomendação colaborativa depende da disponibilidade de uma massa crítica de utilizadores.

Vários métodos têm sido estudados para resolver a dispersão de dados, mas o problema continua a persistir (Anandhan et al. 2018). Apesar destes estudos, a dispersão é muito comum pela pequena porção de avaliações consoante os itens disponíveis (Recommender Systems Handbook 2015). A dispersão das avaliações continua a ser um impedimento significativo no funcionamento dos sistemas de recomendação (Aggarwal 2016).

Conforme Adomavicius e Tuzhilin (2005), para diminuir a dispersão nas avaliações é usada a informação do perfil do utilizador, tanto pelos seus gostos como demograficamente (filtragem demográfica), de modo a calcular a semelhança entre este e outros usufrutuários.

2.3.3 Mudança constante de preferências

Acontece quando um utilizador está em constante mudança de preferências e necessidades.

Manter-se de acordo com as preferências atuais do utilizador e acompanhar as preferências de longo prazo dos utilizadores para que o sistema recomende itens que coincidam com o histórico, é a solução que Khusro, Ali e Ullah (2016) dispõem.

2.3.4 Escala

O problema de escala ocorre quando é difícil processar a enorme quantidade de dados existentes nos dias de hoje (Aggarwal 2016; Khusro, Ali e Ullah 2016).

Khusro, Ali e Ullah (2016) apresentam como conjunto de técnicas o *clustering* (agrupamento) para pesquisa de utilizadores em grupos pequenos em vez de fazer essa pesquisa na base de dados por inteiro e a posterior combinação desse agrupamento e análise de conteúdo com

algoritmos de filtragem colaborativa.

2.3.5 Gray Sheep

Tendo em conta os autores Khusro, Ali e Ullah (2016) as opiniões de um utilizador, neste caso, são díspares das de um grupo e é impossível ganhar benefício nas recomendações.

A integração das técnicas de filtragem colaborativa e filtragem baseada em conteúdo podem dar sugestões de resolução. Os utilizadores podem ser identificados e separados de outros utilizadores utilizando técnicas de *clustering* (por exemplo, *k-means*).

2.3.6 Privacidade

A informação pessoal sobre os utilizadores melhora a assertividade e precisão das recomendações que são feitas. No entanto, a disponibilidade destes dados pode levar à falta de privacidade e insegurança por parte do utilizador em questão (Khusro, Ali e Ullah 2016). A maior parte dos utilizadores acha importante que as suas preferências fiquem privadas, sendo relutantes com o fornecimento de dados para o sistema (Recommender Systems Handbook 2015).

É sobre o problema da privacidade que Khusro, Ali e Ullah (2016) também se dedica procurando soluções de resolução, que ele diz, poderem ser o uso de mecanismos criptográficos e perturbações aleatórias, para permitir a publicação de dados por parte dos utilizadores sem identificações, e uso de tecnologias de web semântica (ontologias e programação neurolinguística - NLP) para mitigar a exposição de informação não pretendida por parte do utilizador. Acrescenta-se ainda *k-identity* e algoritmos de sensibilidade para prevenir violações de privacidade (Recommender Systems Handbook 2015).

2.3.7 Ataques de shilling

Um utilizador malicioso ou algum competidor do item pode entrar no sistema e fazer avaliações falsas (tanto pejorativas como apreciativas), diminuindo o desempenho e a qualidade das recomendações. Quem o afirma são Khusro, Ali e Ullah (2016) que adiantam que estes ataques podem ser detetados através de diferentes maneiras como atributos específicos genéricos e modelados e mudanças de previsão. Os autores também categorizam os ataques por dimensões, tais como o seu tamanho, o conhecimento requerido para iniciar o ataque e as intenções do atacante. Aggarwal (2016) acrescenta que ao criar um conjunto de feedbacks falsos de vários utilizadores diferentes, é possível a mudança de previsões no Sistema.

No que diz respeito aos ataques de *shilling*, Aggarwal (2016) explica que para descobrir os autores destes ataques têm sido feitos alguns métodos que permitam sistemas de recomendação mais robustos. Podem ser usados os *captchas* (*completely automated public turing test to tell computers and humans apart*) que diferenciam humanos e máquinas no contexto da interação *web*. Os *captchas* podem permitir entrar das avaliações, especialmente quando existe um número grande destas com o mesmo ip (Protocolo de internet).

O autor faz referência também a um algoritmo que permite construir sistemas de recomendação confiáveis, onde cada indivíduo é pesado no seu índice de reputação ao fazer a avaliação/recomendação. O mesmo autor adianta que grupos de utilizadores que têm avaliado muitos itens e dão avaliações atípicas consoante as outras avaliações feitas na base de dados, em princípio, serão perfis falsos.

2.3.8 Limitações da análise de conteúdo

É descrito como um sistema que apenas tem uma quantidade limitada de informação dos utilizadores e do conteúdo dos itens (Recommender Systems Handbook 2015).

Adomavicius e Tuzhilin (2005) afirmam que se dois itens diferentes forem representados com as mesmas características, são indistinguíveis, mostrando que as técnicas dos sistemas de recomendação são limitadas pelas características. Este problema refere-se à filtragem baseada em conteúdo.

A qualidade das recomendações pode ser menor devido à dificuldade em extrair informação fidedigna automaticamente dos vários conteúdos como imagens, áudio, texto, diz Bobadilla et al. (2013)

Serão necessárias ontologias (Recommender Systems Handbook 2015), mas poucos autores encontram soluções relativas a este assunto.

2.3.9 Falta de consciência do contexto

De modo a dar recomendações, o sistema terá de estar consciente da situação e dos objetivos do utilizador. Isto é, por exemplo, deve estar atento ao que o utilizador está a fazer (comer, viajar), condições em que o utilizador se encontra, bem como o contexto e época (Recommender Systems Handbook 2015). Este contexto é adquirido nas informações que o utilizador fornece para as redes sociais, como é o caso da localização geográfica que o utilizador dá nas recomendações (Bobadilla et al. 2013). Ou seja, os mesmos autores afirmam que o contexto se foca na informação contextual adicional como tempo, localização e sensores de rede *wireless*.

Há também vários fatores que podem ser pensados, de modo às recomendações serem diretas para o utilizador, tal como o humor deste, e outros parâmetros que podem ser atendidos (Madadipouya e Chelliah 2017; Recommender Systems Handbook 2015).

Madadipouya e Chelliah (2017) apoiam em enriquecer a consciência do contexto em que o utilizador está inserido. Deste modo, para aumentar significativamente a qualidade das recomendações, deve-se usar *Location Based Services* (LBS) que ajuda a localizar as pessoas e suas atividades. Ou seja, por exemplo, se uma pessoa se encontrar numa cidade, o sistema irá recomendar restaurantes dessa mesma cidade.

2.3.10 Especialização excessiva

Este problema ainda foi pouco estudado pelos vários autores. No entanto, Madadipouya e

Chelliah (2017) afirmam que, em alguns sistemas, um utilizador recebe novas recomendações que já foram vistas por ele anteriormente.

Para resolver este problema, é possível introduzir alguma aleatoriedade no sistema, de forma a surgirem outras recomendações diferentes, mas que façam sentido (Adomavicius e Alexander Tuzhilin 2005).

2.4. Medidas de semelhança entre utilizadores

As medidas de semelhança atuam entre duas instâncias, os pares de utilizadores ou os pares de itens (Bobadilla et al. 2013). Segundo os mesmos autores, as medidas mais utilizadas são: Coeficiente de Correlação de *Pearson*; Cosseno; Cosseno ajustado; Distância Euclidiana; *k-Nearest Neighbor*. Estas medidas têm como objetivo aumentar a confiança, reputação e credibilidade das recomendações.

Apesar de, como já foi referido, existirem medidas de semelhança entre itens baseado em memória, as medidas utilizadas são de filtragem colaborativa baseada em memória pela semelhança entre utilizadores.

Este tipo de semelhança inclui 3 etapas (Huang e Gartner 2015):

- Construir perfil do utilizador;
- Calcular semelhanças;
- Agregar em top S mais semelhantes.

2.4.1 kNN

De acordo com Wang et al. (2017), na filtragem colaborativa, o *kNN* depende de uma simples intuição de que um utilizador ativo pode ter comportamentos de preferência em algum item se este item for apreciado por um conjunto de utilizadores semelhantes (vizinhos).

Este algoritmo possui três tarefas principais (Bobadilla et al. 2013):

- Determinar *k* utilizadores “vizinhos” do utilizador ativo ‘*u*’;
- Implementar uma abordagem agregada às avaliações da “vizinhança” em itens não avaliados pelo utilizador ativo ‘*u*’, abordagem esta que pode ser a média e a soma do peso.
- Extrair as previsões do passo 2, depois selecionar as recomendações *top S*. Para obter o *top S* de recomendações, é necessário escolher *k* itens que dão mais satisfação ao utilizador ativo segundo as suas previsões.

É, desta forma, que o “vizinho mais próximo” é calculado pela semelhança, maioritariamente, entre item-item, mas também pode ser usado para o par utilizador-utilizador.

2.4.2 Correlação de Pearson

A medida coeficiente de correlação de Pearson avalia quanto dois utilizadores estão correlacionados. É a medida que captura equivalência entre vetores de avaliações de dois utilizadores e permite medir o grau de correlação entre duas variáveis (Cazella et al. 2009). Os valores variam entre -1 e +1, sendo que, ou existe ausência de correlação ou forte correlação entre as mesmas, respetivamente. O item com avaliação alta prevista é o recomendado (Aggarwal 2015). O coeficiente de *Pearson* é usado para procurar equivalência entre dois utilizadores baseado nos itens avaliados por esses utilizadores (Madadipouya e Chelliah 2017).

Esta medida é calculada da seguinte forma:

$$Pearson = \frac{\sum_{i=1}^s (a1_i - \widehat{a1}) \cdot (u_i - \widehat{u})}{\sqrt{\sum_{i=1}^s (a1_i - \widehat{a1})^2} \cdot \sqrt{\sum_{i=1}^s (u_i - \widehat{u})^2}}$$

Equação 2.1 - Medida Correlação de Pearson (Aggarwal 2015, pág. 607)

Em que \bar{a}, \bar{u} representam as avaliações em comum entre o par de utilizadores, em que um desses utilizadores é o ativo.

2.4.3 Cosseno/Cosseno ajustado

O cosseno normalmente é aplicado em colunas para determinar a semelhança entre as transações, mas pode ser usado também entre itens (Recommender Systems Handbook 2015). Em texto, por exemplo, as funções de semelhança como a medida de cosseno tendem a enfatizar o efeito cumulativo de correspondências em muitos valores de atributos, em vez de grandes distâncias ao longo de atributos individuais. Estes termos gerais também podem ser usados para dados quantitativos. A métrica do cosseno calcula a semelhança entre dois vetores dimensionais baseado no ângulo entre ambos. Esta medida é calculada numa escala de -1 a +1, onde o -1 implica que os objetos são completamente diferentes e o +1 implica que os objetos são completamente semelhantes e o 0 significa que eles não têm nenhuma relação um com o outro.

O cosseno é medido pela seguinte fórmula:

$$sim_{cosseno}(a_1, a_2) = \frac{\sum_i a1_i a2_i}{\sqrt{\sum_i a2_i^2} \sqrt{\sum_i a1_i^2}}$$

Equação 2.2 - Medida Cosseno (Li et al. 2008, equação 7)

2.4.4 Medida simples de semelhança entre utilizadores (Simple_USim)

Esta fórmula compara os POIs semelhantes entre dois utilizadores.

A medida é descrita pelas seguintes etapas (Huang e Gartner 2015):

1. Identificar utilizadores cujo próximo POI após visitar p não foi visitado pelo utilizador ativo;
2. Para os resultados do passo 1, identificar o top S utilizadores mais semelhantes. A medida é utilizada nesta etapa;
3. Para o top S utilizadores mais semelhantes, agregar todos os POIs visitados depois de p pelos utilizadores semelhantes (considerar valor de semelhança entre utilizadores);
4. Selecionar o POI com valor previsto mais alto e recomendar ao utilizador ativo.

Esta fórmula é calculada da seguinte forma:

$$simple_{USim(a,b)} = \frac{|POIS_{a1,b2}|}{\sqrt{|POIS_{a1}| * |POIS_{b2}|}}$$

Equação 2.3 - Medida Simple_USim (Huang e Gartner 2015, pág. 7)

POISa e POISb são o conjunto de POIs visitados pelo utilizador a e o utilizador b, respetivamente. POISa,b é o conjunto de POIs que foram visitados ambos pelo utilizador a e pelo utilizador b. |POISa| é o tamanho (número de elementos) do POISa.

2.4.5 Distância euclidiana

A distância euclidiana é a função mais comumente usada, em comparação a outra variedade de funções existentes. É um tipo de métrica de semelhança para as recomendações que usam avaliações e informações pessoais como *input* (Recommender Systems Handbook 2015). É a distância entre dois pontos de dados. É definida num caminho de dados multidimensionais onde o comportamento dos valores dos atributos são interpretados como dimensões. Quanto mais altos os valores da métrica, maior a distância e a essa distância é menor quanto mais baixo forem os valores.

2.4.6 Proximidade-Impacto-Popularidade

Proximidade-Impacto-Popularidade (PIP) é a medida mais recente que pretende combater os problemas que as medidas tradicionais podem acarretar (Recommender Systems Handbook 2015).

A medida PIP é usada na filtragem colaborativa na semelhança entre utilizadores. É uma medida de semelhança heurística que considera três aspetos: a proximidade, o impacto e a popularidade (Sheugh e Alizadeh 2015). É utilizada para resolver os problemas do cosseno e da correlação de pearson, no sentido de melhorar os defeitos da filtragem colaborativa, como o *cold-start* ou a dispersão de dados.

O autor Ahn (2008) sugere que esta medida seja combinada com outras medidas de semelhança, de modo a aumentar os bons resultados.

Calcula-se da seguinte forma:

$$SIM(a1, a2) = \sum_{k \in Ca1, a2} PIP(r_{a1k}, r_{a2k})$$

Equação 2.4 - Medida PIP (Ahn 2008, pág. 42)

2.4.7 Semelhança Máxima de Padrões de Trajetórias Semânticas

Esta medida assemelha apenas utilizadores e não itens. Para fazer a filtragem colaborativa por trajetos, inicialmente temos de encontrar utilizadores com trajetos semelhantes aos do utilizador ativo.

Deste modo, Ying et al. (2010), descrevem a medida *Maximal Semantic Trajectory Pattern Similarity (MSTPSimilarity)* como a mais indicada para fazer a comparação entre trajetos percorridos por dois utilizadores diferentes, analisando o comportamento e movimento de cada um. Este cálculo é feito de modo a recomendar um potencial caminho novo para o utilizador ativo. O MSTPS é calculado a partir da média de índices comuns entre eles.

Inicialmente, deve-se fazer a extração de todos os locais que o utilizador visita, extraíndo o trajeto inteiro. Depois é utilizada a mesma medida no sentido de comparar a semelhança entre dois trajetos. E por último, deve ser feito um agrupamento, em que os utilizadores com trajetos semelhantes são postos no mesmo grupo.

Após estes passos, o próximo local pode ser previsto não apenas do histórico do utilizador ativo mas também pelo histórico de todos os utilizadores.

Os autores Ying et al. (2012) acrescentam ainda que se um padrão for descoberto de um conjunto de trajetos de um utilizador, podemos prever que esse utilizador, por exemplo, após ir ao parque costuma ir à escola.

Para fazer esta semelhança entre dois padrões de trajetos, calcula-se o LCS (*Longest Common Sequence*), para saber a sequência mais comum entre dois trajetos feitos por dois utilizadores.

Como foi referido, tendo N e Q como dois utilizadores, em que N = “escola, cinema, parque ou banco, restaurante” e q= “escola ou mercado, parque, restaurante”, o seu LCS (N,Q) = Escola, parque, restaurante”.

A figura mostra como é calculado o padrão mais comum de sequência.

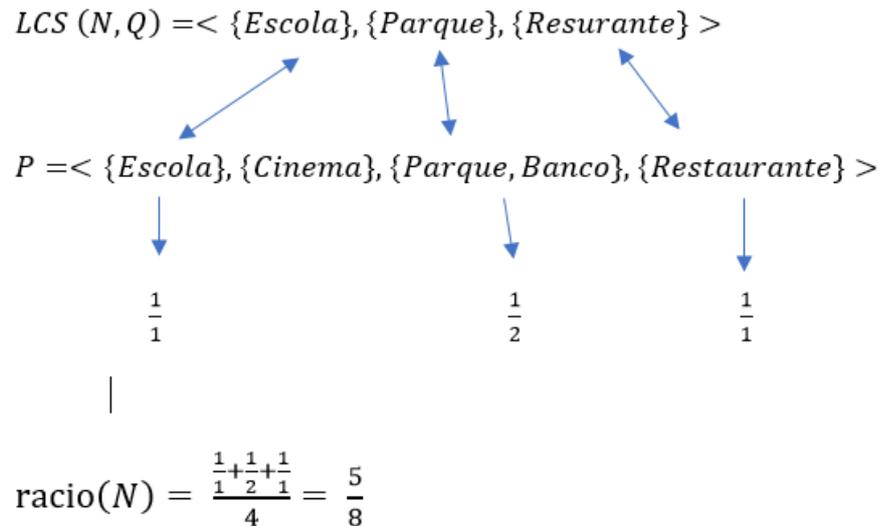


Figura 2.3 – Exemplo de trajetórias semânticas (Ying et al. 2011, figura 9)

Depois é feito um agrupamento com os utilizadores mais comuns. Os autores afirmam que dois utilizadores são mais semelhantes quando têm mais partes em comum, isto é, neste caso, quando têm mais POIs em comum do trajeto que percorreram.

Após saber o rácio dos dois padrões, pode-se calcular a semelhança máxima de padrão entre trajetos pela média de rácios das partes comuns entre os utilizadores.

Ou seja, por um lado, pode-se calcular a média igual, representada na fórmula seguinte:

$$MSTP - Similarity_{EA}(N, Q) = \frac{racio(LCS(N, Q), N) + racio(LCS(N, Q), Q)}{2}$$

Equação 1.5 - Medida MSTPS pela média de rácios das partes comuns entre utilizadores (Ying et al. 2010, equação 3)

Por outro lado, podemos calcular a média de peso, em proporção ao tamanho dos padrões, como mostra a seguinte formula:

$$MSTP - Similarity_{WA}(N, Q) = \frac{|N| \times racio(LCS(N, Q), N) + |Q| \times racio(LCS(N, Q), Q)}{|N| + |Q|}$$

Equação 2.6 - Medida MSTPS pela média de peso de cada trajeto (Ying et al. 2010, equação 4)

Um padrão longo fornece mais informação sobre o comportamento dos utilizadores do que um padrão curto. Assim um padrão longo dá mais peso na medida de semelhança entre dois padrões (Ying et al. 2011).

2.5. Medidas de Avaliação em Sistemas de Recomendação

Os sistemas de recomendação usam um número de diferentes medidas para avaliar o sucesso da recomendação ou algoritmos de previsão. Um tema recorrente nas pesquisas de filtragem colaborativa é gerir métricas para calcular com precisão, rigor e exatidão, a existência de semelhanças entre utilizadores ou itens. As métricas têm sido desenhadas para se adaptarem às características, peculiaridades e restrições dos SR.

Segundo Bobadilla et al. (2013) as medidas de avaliação são divididas em quatro tipos: avaliação das previsões, que conferem precisão, ou qualidade destas; Conjunto de métricas de recomendação que conferem qualidade; Qualidade da lista de recomendações; Métricas de diversidade.

2.5.1 Avaliação das previsões, que conferem precisão, ou qualidade destas

As medidas de precisão medem quão bem o sistema pode prever um valor exato de avaliação para um item específico (Herlocker et al. 2004).

Estas são medidas para avaliação de problemas de regressão, que avaliam o erro entre as avaliações previstas pelo sistema e a matriz das avaliações utilizada pelo sistema para gerar a previsão.

Deve-se usar *Mean Absolute Error* (MAE) e *Root Mean Squared Error* (RMSE) separadamente para cada item e calcular a média de todos os itens, como medidas representantes de cada erro previsto de cada item. De RMSE e MAE dependem apenas da magnitude dos erros cometidos. RMSE talvez seja de entre as métricas usadas, a mais popular na avaliação da precisão das avaliações previstas (Recommender Systems Handbook 2015). Segundo Cazella et al. (2009) a métrica MAE permite calcular o desvio absoluto médio entre as previsões do Sistema de Recomendação e as avaliações feitas pelo utilizador. A diferença obtida é vista como erro de previsão.

Cobertura é a capacidade de prever a partir de uma métrica, aplicado a um SR específico. Calcula a percentagem de situações onde pelo menos o “vizinho mais próximo” de cada utilizador ativo pode avaliar um item que não foi avaliado até agora por esse mesmo utilizador (Recommender Systems Handbook 2015).

O cálculo das métricas de erro pode ser feito a partir da comparação das avaliações atribuídas por um utilizador a um item com a avaliação que o sistema prevê que seja essa classificação atribuída por esse utilizador ao item que se pretende recomendar.

2.5.2 Conjunto de métricas de recomendação que conferem qualidade

Precisão e sensibilidade (*recall*) são medidas de desempenho usadas maioritariamente na

recuperação da informação, sendo de avaliação para problemas de classificação.

A precisão mede a eficiência dos itens selecionados relevantes; sensibilidade mede o número de itens selecionados relevantes e o número total de itens em todo o conjunto. Precisão mede quão bem as avaliações dos utilizadores podem ser reproduzidas (Anandhan et al. 2018); sensibilidade é usada para especificar a percentagem de itens relevantes que foram recuperados por um sistema de busca.

Estas técnicas podem ser usadas na avaliação de algoritmos de recomendação, no entanto não são suficientes para a satisfação do utilizador (Recommender Systems Handbook 2015).

Muitos estudos apontam para uma combinação entre as duas métricas, combinando numa única, chamada F1 (Herlocker et al. 2004).

2.5.3 Qualidade da lista de recomendações

Segundo Bobadilla et al. (2013) esta medida de classificação da qualidade refere que quando há uma quantidade enorme de dados, os utilizadores vão dar preferência aos primeiros itens da lista de recomendações.

A medida DCG tem um conjunto de vantagens (Järvelin e Kekäläinen 2017):

- Reduz o ganho recebido através de documentos encontrados posteriormente nos resultados classificados;
- Permite modelar a persistência do utilizador ao examinar listas de resultados de classificação longa ajustando o fator de desconto.

Esta medida combina simultaneamente avaliação do documento (baseado nas probabilidades de relevância) e grau de relevância dos documentos.

2.5.4 Métricas de diversidade

Bobadilla et al. (2013) descrevem as métricas “novidade” e diversidade.

A métrica “novidade” é o grau de diferença entre os itens recomendados para cada utilizador e conhecidos por ele. Ocorre quando o sistema sugere a um utilizador um item não conhecido que ele pode ter descoberto autonomamente. Os autores Castells, Vargas e Wang (2015) acrescentam que a novidade alude, geralmente a quão diferente a informação é ao que já foi visto anteriormente.

Diversidade é o grau de diferenciação entre os Sistemas de Recomendação. Permite a possibilidade de o utilizador encontrar o item correto. Esta diversidade aplica-se a um conjunto de itens e está relacionada com a diferença entre os itens e nas suas relações (Castells, Vargas e Wang 2015).

2.6. Implementação de um Sistema de Recomendação na aplicação do Museu Digital da Universidade do Porto

A aplicação #IWASHERE contém informações sobre roteiros e seus respectivos POIs bem como a localização geográfica, como representam as figuras seguintes:

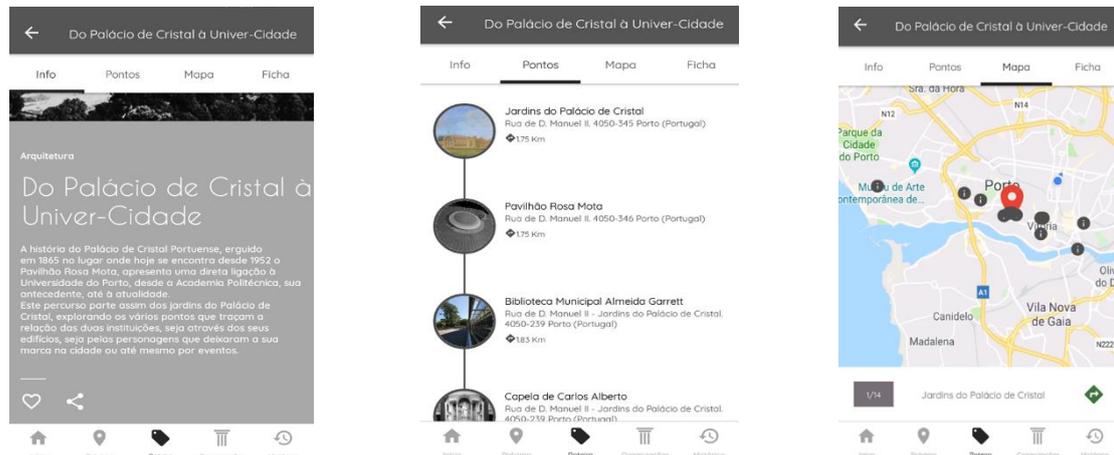


Figura 2.3 - Roteiro com pontos de interesse e suas respectivas localizações geográficas

A visita à cidade do Porto pode acarretar alguns contratempos associados à quantidade vasta de informação, ao tamanho da cidade, ao tempo disponível da pessoa e à qualidade daquilo que se procura e encontra. Planejar uma viagem ou uma visita a determinados locais pode ser muito moroso, principalmente quando queremos que as nossas necessidades sejam satisfeitas. Este planeamento, normalmente, envolve a identificação dos pontos que temos interesse em visitar, bem como a percepção da média do tempo que vamos dispensar de um local para o outro e o tempo total que gastaremos na visita.

A criação de um Sistema de Recomendação propõe solucionar estas dificuldades, e ajudar o visitante na sua viagem. Segundo Solima, Peruta e Maggioni (2016), este sistema usa as preferências desse mesmo utilizador, expressas pelo comportamento ou movimentos durante a visita.

Para melhorar a vivência do utilizador, inicialmente, é necessário entender detalhadamente as suas escolhas, envolvendo a identificação de mecanismos que permitam que as informações sejam filtradas de maneira a os adaptar ao mesmo.

Ao implementar um itinerário de visita, é possível criar Pontos de Interesse (POIs) de cada utilizador. Assim, os visitantes ficam em condições de ver o que realmente necessitam, sendo que a principal preocupação é que não haja repetição do local de exibição.

Solima, Peruta e Maggioni (2016) reconhecem que, deste modo, o utilizador tira vantagens desta criação pois, para além do aumento do prazer durante a sua visita, este consegue visitar

realmente tudo aquilo que necessita, com um percurso específico, num período de tempo menor. A complexidade associada à vasta informação disponível pode ser decomposta e traduzida em informação digital estruturada, usando estes métodos tecnológicos que melhoram, supervisionam e processam o comportamento do visitante.

O SR tem, então, um papel importante na tomada de decisão do utilizador na escolha de um itinerário, na decisão do tempo que vai gastar em cada local que pretende visitar e na sequência de lugares a visitar.

Todos os utilizadores que façam uma visita individual ou em grupo e até mesmo os turistas beneficiam desta implementação. Estes últimos benfeitorizam pela oportunidade de conhecer melhor a cidade e pontos turísticos que possam ter interesse dentro dos caminhos sugeridos. A universidade do Porto e seus museus podem representar uma mais-valia na expansão turística do Porto, consoante a qualidade histórica que os museus possuem (Pinto et al. 2016). É assim fundamental referir que o património universitário e o turismo podem representar uma vantagem significativa na divulgação da cidade do Porto.

2.7 – Complicações ao longo do projeto

Inicialmente, o objetivo principal desta dissertação centrou-se na criação de um Sistema de Recomendação, em que, a abordagem metodológica mostrava as etapas desta criação. Isto é, seria essencial obter dados das visitas por partes dos utilizadores, e com isto, com a ajuda da empresa Weblevel, seria programado, para posteriormente utilizar o software de gestão RapidMiner para fazer a criação a partir de uma filtragem colaborativa.

A falta de dados neste projeto foi a complicação maior, não permitindo a criação do Sistema de Recomendação, o que se tentou resolver por outra via, de modo a dar todas as ferramentas necessárias para serem utilizadas num projeto futuro.

Consequentemente, foi difícil a implementação ou experimentação no trabalho de desenvolvimento.

Outra dificuldade deteve-se na falta de estudos que se referiram a recomendações de percursos completos.

3. Metodologia

Este projeto atua no sentido de tornar mais fácil a visita de vários pontos de interesse na cidade do Porto, sendo possível pela criação da aplicação #IWASHERE. Esta aplicação, como anteriormente descrito, contém vários roteiros e pontos de interesse da cidade do Porto, desafiando elucidar os seus utilizadores na visita desses determinados locais.

Para tal, inicialmente é necessário clarificar que um roteiro é um conjunto de dados geográficos que capturam o comportamento físico de movimentos dos utilizadores (Ying et al. 2010). Ou seja, um trajeto pode ser definido como um conjunto de pontos = P1 -> P2 -> Pn, em que cada ponto consiste num conjunto de coordenadas. A aplicação contém esta informação em que esses pontos são localizações a serem visitadas na cidade do Porto, onde o utilizador simplesmente poderá escolher se visita ou não esse local.

As rotas, que são percursos sugeridos pela plataforma, podem constituir um ou mais pontos de interesse, ou seja, dentro de uma rota pode ter apenas um POI (por exemplo a Igreja dos Clérigos), e outra rota existente pode incluir dois ou mais POIs (por exemplo Palácio de Cristal que tem os jardins ou o Pavilhão Rosa Mota).

Para a realização deste projeto é necessário recolher informações sobre as visitas dos utilizadores, satisfazendo os seguintes atributos:

- IDutilizador: código identificador de cada utilizador;
- IDcaminho: código identificador de cada caminho percorrido;
- Caminho – lista ordenada de Pontos de Interesse visitados;
- IDTempo – tempo de deslocação entre cada POI e tempo total da visita.

O tempo de paragem num local apenas é importante para definir se é um local de interesse para o utilizador.

Cada rota deve então ter um identificador, com os POIs associados a cada uma. Aquando a visita do POI, o utilizador irá fornecer um ID que o diferenciara de todos os outros utilizadores. Deverá saber-se quais POIs foram visitados, quanto tempo cada utilizador demorou durante a visita de um POI a outro e qual o tempo total da visita do trajeto completo.

Esta metodologia permitirá demonstrar todas as ferramentas no sentido de fazer uma recomendação ao utilizador ativo. É indispensável usar a filtragem colaborativa, uma das técnicas mais utilizadas para fazer recomendações.

Como já referido, a não existência de dados dificultou-nos a implementação do Sistema de Recomendação.

3.1. Porquê utilizar a Filtragem Colaborativa?

Burke (2002) considera a filtragem colaborativa como a tecnologia mais madura e afirma que é

a técnica mais popular e largamente usada em Sistemas de Recomendação.

Como vimos anteriormente, Batet et al. (2012) afirmam que este tipo de filtragem procura utilizadores semelhantes e seleciona esses utilizadores segundo a sua proximidade.

Vários estudos mostram que o uso desta filtragem é o mais adequado e com sucesso, sendo utilizada na amazon, no netflix e no spotify (RS handbook 2015).

Para este trabalho é necessário fazer a semelhança entre utilizadores no sentido de comparar trajetos percorridos ou POIs visitados. É aqui que a filtragem colaborativa atua pois ela calcula semelhanças na história de avaliações dos utilizadores (RS handbook 2015). Assim, pretende prever o que irá ser recomendado ao utilizador ativo a partir do histórico dos utilizadores.

Para além disso, a FC estuda o comportamento do utilizador durante a visita. Este comportamento suporta variáveis, sabendo se o utilizador visita ou não um POI, o tempo que demora em cada POI e o tempo que demora em toda a visita. As atitudes que o utilizador tem durante a visita são um fator importante no projeto em questão pois permite otimizar o problema e facilitar a recomendação.

Apesar da filtragem baseada em conteúdo ser também muito utilizada em Sistemas de Recomendação, ela apenas estuda as características de cada item/POI, e não as características de cada utilizador.

Como o problema requer calcular a semelhança entre utilizadores, utilizando o histórico de cada e estudando o seu comportamento, será mais adequado utilizar a filtragem colaborativa. É importante referir que esta pretende calcular padrões, sendo estes padrões o trajeto, os locais ou sequência de pontos que o utilizador pode visitar.

3.2. Dois panoramas diferentes

Para este trabalho são defendidas duas vias para propor trajetos novos com POIs que o utilizador ativo ainda não tenha visitado.

Foi necessário então, definir as diferentes medidas de semelhanças entre os utilizadores, de modo a fazer a comparação entre eles consoante os trajetos percorridos e os pontos de interesse visitados.

Os passos definidos encontram-se na tabela 2.2.

Tabela 2.2 – Uso da Filtragem Colaborativa em semelhanças de trajetos e de POIs

Filtragem Colaborativa por trajetos	Filtragem colaborativa por POIs
1.1 – Encontrar utilizadores com trajetos semelhantes.	2.1 – Encontrar utilizadores com maior número de POIs visitados que o utilizador ativo também tenha visitado.
1.2 – Selecionar os top S utilizadores de acordo com a semelhança dos trajetos.	2.2 – Selecionar os top S utilizadores de acordo com a semelhança entre POIs visitados.
1.3 - Propor trajetos usando POIs visitados pelos utilizadores semelhantes que o utilizador ativo não tenha visitado.	2.3 - Propor trajetos novos usando POIs visitados pelos utilizadores semelhantes que o utilizador ativo não tenha visitado.

3.2.1 Filtragem colaborativa por trajetos

Muitos estudos ainda são feitos a partir de semelhanças entre características geográficas, mas é necessário haver estudos com características semânticas, pois apenas com características geográficas não são suficientes. Esta semelhança é feita com base na consciência do que é trajetória semântica, em que é uma sequência de localizações com *tags* semânticos para capturar pontos de referências por onde passou o utilizador, e esta semântica é diferente da localização geográfica e da proximidade dos POIs pela localização.

Outros autores como Lee e Han (2007) apresentam um outro método para calcular a semelhança entre duas trajetórias, *Partition and Group*. No método ele procura pontos característicos entre trajetos num processo de segmentação e aplica três medidas de distância entre trajetos. Não se adequa ao problema porque este método usa estas distâncias aplicadas à informação geográfica e não pode ser usado para medir semelhança entre utilizadores numa trajetória semântica.

Para calcular a semelhança entre trajetos verificou-se que a medida mais adequada é o *MSTPSimilarity (Maximal Semantic Trajectory Pattern Similarity)*, pois esta medida compara dois trajetos inteiros percorridos por dois utilizadores diferentes.

3.2.2 Filtragem colaborativa de Pontos de Interesse

Neste ponto são apresentadas medidas que fazem a semelhança entre utilizadores a partir do histórico dos pontos de interesse visitados e não a partir do trajeto completo realizado.

Para este efeito, existe um conjunto de medidas que permitem fazer esta comparação, ajudando a perceber a “vizinhança” mais próxima do utilizador ativo. Pode ser usada a Correlação de Pearson, no entanto o cosseno é a medida que mostra melhores resultados empíricos (Aggarwal 2016). Os níveis de precisão de cada uma das medidas são muito idênticos, no entanto a correlação de Pearson é particularmente apropriada e eficaz quando queremos um número

grande de previsões com qualidade (Sánchez et al. 2008). Ahn (2008) afirma que a correlação de *Pearson* mostra a melhor performance, no entanto, a medida PIP produz o melhor resultado. Esta medida é mais recente que supera os problemas do *Cold-Start* que as medidas tradicionais ainda não travaram.

Deste modo, e referente ao trabalho proposto, visto todas as medidas apresentadas serem de semelhança entre utilizadores, foram escolhidas as mais importantes para representar o projeto. Além do cosseno e da correlação de *Pearson*, a medida PIP e a medida *Simple_USim* são as mais apropriadas.

3.2.3 Pseudocódigo

A formulação do pseudocódigo é escrito numa linguagem simples e é apresentado para facilitar a leitura do código. A não implementação deve-se aos dados que a empresa Weblevel não conseguiu receber e fornecer, bem como, conseqüentemente, ao tempo para a empresa o implementar. No entanto, foi importante, para cada um dos cenários presentes, formular os seus pseudocódigos. Nos anexos são apresentados os pseudocódigos para Filtragem Colaborativa por trajetos e para Filtragem Colaborativa por Pontos de Interesse. É mostrado também um argumento de maximização que pretende propor trajetos usando Pontos de Interesse visitados pelos utilizadores semelhantes ao utilizador ativo, os quais este ainda não tenha visitado.

3.3. Otimização do problema

De forma a encontrar uma solução ótima para o problema, foi decidido implementar uma solução de otimização, de modo a resultar no melhor desempenho possível do sistema.

Para isso, foi estudada a programação linear de forma a otimizar uma função de variáveis sujeita a uma série de restrições.

Como existem vários POIs que o utilizador percorre num trajeto, o objetivo será então recomendar o maior número de POIs possíveis em menor tempo, de modo a satisfazer as necessidades do utilizador.

É importante que esta otimização faça com que todos os utilizadores passem por vários POIs uma única vez, sem que haja repetição na paragem.

Para este problema, é necessário responder a um conjunto de etapas de modo a fazer a sua tradução.

1. Apresentar as variáveis do problema.

As variáveis são as características dos elementos da amostra que nos interessa averiguar.

São elas:

- O tempo total do percurso;

- O número de POIs visitados;
- A distância entre cada POI;
- Se o utilizador visitou ou não um determinado POI.

Todas as variáveis têm de ser discretas, isto é, números inteiros.

2. Descrever o tipo de problema.

No caso proposto é pretendido maximizar a função, ou seja, aumentar o resultado da função de otimização que, neste caso, é multiobjetivo. Assim, todas as variáveis apresentadas devem ser maximizadas, sujeitas a restrições. Consequentemente, a satisfação do utilizador aumenta, pela precisão e veracidade das recomendações feitas.

3. Função-objetivo

$$a \sum POIs + b \sum Tempo + c \sum Dist$$

Equação 2.7 - Função-objetivo

Pretende-se maximizar a soma de todos os POIs, do tempo total e da distância entre cada POI.

4. Restrições

Para cada variável existe uma restrição.

Assim, dado um conjunto S de POIs (S=10, por exemplo), o número de POIs a ser visitado por cada utilizador terá de ser, no máximo, 8.

O tempo máximo que o utilizador irá gastar na visita de todos os POIs recomendados é de 3 horas.

A distância máxima entre 2 POIs diferentes será de 1km.

Estas restrições definem como deve ser feita a recomendação ao utilizador ativo, sem que sejam ultrapassados os valores definidos.

5. Condições de não negatividade

É considerado que todas as variáveis tenham valor positivo.

6. Problema traduzido

$$Máx = \sum POIs + \sum Tempo + \sum Dist$$

Equação 2.8 - Problema de otimização traduzido

Sujeito a:

Máx POIs \leq 10 (Restrição 1)

Máx Tempo \leq 3 horas (Restrição 2)

Máx Dist \leq 1km (Restrição 3)

0 \leq a, b, c \leq 1

A + b + c = 1

POIs, Tempo, Dist \geq 0

3.4. Contribuições do projeto

O *cold-start* é um intrave nos sistemas de recomendação que precisa de mais estudos e integração por pesquisas e trabalhos futuros. A filtragem colaborativa sofre deste problema e tenta procurar soluções que envolvem analisar metadados ou fazer perguntas iniciais aos novos utilizadores de modo a saber as suas preferências iniciais.

Apesar dos vários estudos referidos, o projeto incidiu noutras soluções que foram vistas por outros autores. Deste modo, a escolha da utilização da medida PIP, como já vimos, destina-se a reduzir os problemas da FC.

Ao otimizar o problema definimos um conjunto de restrições das variáveis, sendo que o novo utilizador no sistema terá de cumprir essas mesmas restrições. Sendo assim, o sistema já saberá o máximo de tempo que o utilizador pode dispensar e que máximo de POIs ele poderá ver. É também definido a distância máxima que dois POIs devem ter entre si, não cansando o utilizador que faz o trajeto.

Quando um utilizador novo entra no sistema, este já dispõe ao seu encargo uma amostra de roteiros com determinados POIs, o que facilitará a visita deste novo utilizador pois pode pesquisar que roteiros tem disponíveis. Após percorrer um POI, o sistema já determinará que POIs ou roteiros o utilizador deseja fazer.

De modo a guardar os dados do utilizador referentes ao seu histórico, a aplicação deverá estar ligada a um servidor que guardará essa informação. Assim, é possível saber os utilizadores semelhantes protegendo a privacidade de cada um.

4. Conclusões e perspectivas de desenvolvimento

Visto ser objetivo do MDUP ajudar na crescente procura do património científico, museológico, arquitetónico, etc., a criação da aplicação #IWASHERE veio de encontro a esta finalidade, apresentando diferentes rotas com determinados pontos de interesse da cidade do Porto. Esta aplicação é pensada exactamente no sentido de filtrar a informação necessária para cada tipo de visitante com diferentes características e gostos. É destinada a todas as pessoas que queiram visitar estes pontos, mas é sobretudo significativo para o turismo. O ponto fulcral desta dissertação recai essencialmente neste sentido, na criação de um Sistema de Recomendação, que permita recomendar esses pontos ou rotas a um determinado utilizador. A importância desta dissertação é notável no sentido de compreender o conjunto de passos necessários e indispensáveis para a criação do Sistema de Recomendação para a aplicação.

Apesar dos vários tipos apresentados, o uso da Filtragem Colaborativa mostrou-se o mais adequado, sendo também o mais utilizado nos estudos referentes a SR.

De modo a criar o SR, é importante fazer a semelhança entre utilizadores, permitindo agrupá-los por gostos semelhantes. Assim, esta semelhança pode ser calculada entre utilizadores do grupo em vez de compará-lo com todos os utilizadores do sistema. As medidas de semelhança entre os utilizadores mostrou que, para além das tradicionais usadas (cosseno e correlação de pearson), existem outras que comportam as mesmas características mas que permitem um aperfeiçoamento das anteriores. Os estudos mais recentes apostam nessas mesmas medidas novas, de modo a tentar resolver os problemas mais recorrentes – *cold-start* e esparsidade de dados – que existem no tipo de filtragem colaborativa.

As medidas de avaliação apresentadas permitem, num passo posterior, avaliar todo o Sistema de Recomendação.

Sendo objetivo propor trajetos usando POIs visitados pelos utilizadores semelhantes que o utilizador ativo não tenha visitado, a otimização do problema torna-se fundamental. Esta otimização, apesar da falta de dados fornecidos, vai permitir identificar as variáveis do problema e fazer restrições para cada uma. Isto permite facilitar o problema inicial, pois, a recomendação de um POI ou trajeto ao utilizador ativo, é feita segundo essas restrições. Permite maximizar o número de POIs que devem ser visitados, maximizar o tempo total da viagem (de modo que o utilizador não fique muitas horas) e maximizar a distância entre os POIs (para que a distância máxima entre dois POIs seja pouca).

4.1. Perspetivas futuras

Apesar das várias medidas de semelhança utilizadas, é importante definir medidas que tenham o mesmo intuito mas que consigam resolver o problema do *Cold-start*. Assim, é importante que

o estudo deste tema seja mais recorrente.

Não obstante a existência bastantes estudos de SR com itens e utilizadores e que provam o sucesso da implementação, poucos são os estudos referem recomendações de percursos completos. Por isso é importante que haja estudos mais afincados sobre Sistemas de Recomendação adaptados a sequências e não apenas segundo avaliações a itens feitas pelos utilizadores.

O uso da *web 3.0*, que inclui dados relacionados com contexto, que recolhem dados através de sensores (temperatura, hábitos alimentares, etc.), deverá ser pensada pois permitirá uma recomendação mais adequada e direcionada a cada utilizador.

No futuro é necessário que, com o apoio desta dissertação se implemente o algoritmo apresentado num *software* de gestão de dados, como é exemplo o *RapidMiner*. A partir desta implementação é necessário utilizar a medida de avaliação mais adequada.

5. Referências Bibliográficas

- Adomavicius, G., e Alexander Tuzhilin. 2005. «Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions». *IEEE Transactions on knowledge and data engineering* 17, no. 6: 734–49. <https://doi.org/10.1109/TKDE.2005.99>.
- Aggarwal, Charu C. 2015. *Data Mining: the textbook*. Springer. New York: Springer. <https://doi.org/10.1007/978-3-319-14142-8>.
- Aggarwal, Charu C. 2016. «Neighborhood-Based Collaborative Filtering». Em *Recommender Systems: the textbook*, editado por Springer International Publishing Switzerland, 29–69. Switzerland. <https://doi.org/10.1007/978-3-319-29659-3>.
- Ahn, Hyung Jun. 2008. «A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem». *Information Sciences* 178, no. 1: 37–51. <https://doi.org/10.1016/j.ins.2007.07.024>.
- Alyari, F., e Nima Jafari Navimipour. 2018. «Recommender systems: A systematic review of the state of the art literature and suggestions for future research». *Kybernetes* 47, no. 5: 985–1017. <https://doi.org/10.1108/K-06-2017-0196>.
- Anandhan, A., Liyana Shuib, Maizatul Akmar Ismail, e Mujtaba Ghulam. 2018. «Social Media Recommender Systems: Review and Open Research Issues». *IEEE Access* 6 (agosto): 15608–28. <https://doi.org/10.1109/ACCESS.2018.2810062>.
- Batet, M., Antonio Moreno, David Sánchez, David Isern, e Aida Valls. 2012. «Turist@: Agent-based personalised recommendation of tourist activities». *Expert Systems with Applications* 39, no. 8: 7319–29. <https://doi.org/10.1016/j.eswa.2012.01.086>.
- Bellogín, A., e Pablo Sánchez. 2017. «Collaborative filtering based on subsequence matching: A new approach». *Information Sciences* 418–419: 432–46. <https://doi.org/10.1016/j.ins.2017.08.016>.
- Binucci, C., Felice De Luca, Emilio Di Giacomo, Giuseppe Liotta, e Fabrizio Montecchiani. 2017. «Designing the Content Analyzer of a Travel Recommender System». *Expert Systems with Applications* 87: 199–208. <https://doi.org/10.1016/j.eswa.2017.06.028>.
- Bobadilla, J., Antonio Hernando, Fernando Ortega, e Jesus Bernal. 2011. «A framework for collaborative filtering recommender systems». *Expert Systems with Applications* 38, no. 12: 14609–23. <https://doi.org/10.1016/j.eswa.2011.05.021>.
- Bobadilla, J., Fernando Ortega, Antonio Hernando, e Abraham Gutiérrez. 2013. «Recommender systems survey». *Knowledge-Based Systems* 46: 109–32. <https://doi.org/10.1016/j.knosys.2013.03.012>.
- Burke, R. 2002. «Hybrid Recommender Systems: Survey and Experiments». *User Modeling and User-Adapted Interaction* 12, no. 4: 331–370.
- Cai, G., Kyungmi Lee, e Ickjai Lee. 2018. «Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos». *Expert Systems with Applications* 94: 32–40. <https://doi.org/10.1016/j.eswa.2017.10.049>.
- Çano, E., e Maurizio Morisio. 2017. «Hybrid recommender systems: A systematic literature review». *Intelligent Data Analysis* 21, no.6: 1487–1524. <https://doi.org/10.3233/IDA-163209>.
- Castells, P., Saúl Vargas, e Jun Wang. 2015. «Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance». Em *International Workshop on Diversity in Document Retrieval (DDR 2011)*. <https://doi.org/10.1007/978-0-387-85820-3>.
- Cazella, S. C., Eliseo Berni Reategui, Munique Machado, e Jorge Luis V. Barbosa. 2009. «Recomendação de Objetos de Aprendizagem Empregando Filtragem Colaborativa e Competências». Em *XX Simpósio Brasileiro de Informática na Educação (SBIE)*. Brasil.

- Guo, G., Zhang Jie, e Daniel Thalmann. 2014. «Merging trust in collaborative filtering to alleviate data sparsity and cold start». *Knowledge-Based Systems* 57: 57–68. <https://doi.org/10.1016/j.knosys.2013.12.007>.
- Herlocker, J. L., Joseph A. Konstan, Loren G. Terveen, e John T. Riedl. 2004. «Evaluating collaborative filtering recommender systems». *ACM Transactions on Information Systems* 22, no. 1: 5–53. <https://doi.org/10.1145/963770.963772>.
- Hernando, A, Jesús Bobadilla, Fernando Ortega, e Abraham Gutiérrez. 2017. «A probabilistic model for recommending to new cold-start non-registered users». *Information Sciences* 376: 216–32. <https://doi.org/10.1016/j.ins.2016.10.009>.
- Huang, H., e Georg Gartner. 2015. «Using trajectories for collaborative filtering-based POI recommendation». *International Journal of Data Mining, Modelling and Management* 6, no. 4: 333. <https://doi.org/10.1504/ijdm.2014.066762>.
- ICOM. 2015. «ICOM Portugal». Definições. 2015. <http://icom-portugal.org/2015/03/19/definicao-museu/>.
- Järvelin, K., e Jaana Kekäläinen. 2017. «IR evaluation methods for retrieving highly relevant documents». Em *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, 51:243–50. Nova York: ACM Digital Library. <https://doi.org/10.1145/3130348.3130374>.
- Khusro, S., Zafar Ali, e Irfan Ullah. 2016. «Recommender Systems: Issues, Challenges, and Research Opportunities». *Information Science and Applications (ICISA)* 376 (fevereiro): 1179–89. https://doi.org/10.1007/978-981-10-0557-2_112.
- Lee, Jae-gil, Jiawei Han, e Kyu-Young Whang. 2007. «Trajectory Clustering: A Partition-and-Group Framework». *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. Beijing, China. <https://doi.org/10.1145/1247480.1247546>.
- Li, Q., Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, e Wei-Ying Ma. 2008. «Mining user similarity based on location history». Em *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, 19–26. California: ACM Digital Library. <https://doi.org/10.1145/1463434.1463477>.
- Madadipouya, K., e Sivananthan Chelliah. 2017. «A Literature Review on Recommender Systems Algorithms, Techniques and Evaluations». *BRAIN. Broad Research in Artificial Intelligence and Neuroscience* 8, no. 2: 109–24. <http://brain.edusoft.ro/index.php/brain/article/view/693>.
- Pinto, M. M., Susana Medina, Rodolfo Matos, e Paulo Fontes. 2016. «U.Openlab Methodology: a Conceptual Model and Flowchart for the Dynamic Co-Production and (Re)Use of Digital Contents». *ICER2016 Proceedings* 1 (novembro 2016): 4812–19. <https://doi.org/10.21125/iceri.2016.2140>.
- Poulot, D. 2013. «O que é um museu?» Em *Museu e Museologia: Autêntica*. Belo Horizonte.
- Ricci, F., Lior Rokach, e Bracha Shapira, eds. 2015. *Recommender Systems Handbook*. Second Edi. Nova York: Springer. <https://doi.org/10.1007/978-1-4899-7637-6>.
- Rolim, V., Rafael Ferreira, Evandro Costa, Anderson Cavalcanti, e Máverick Dionísio. 2017. «Um Estudo Sobre Sistemas de Recomendação de Recursos Educacionais». Em *Anais dos Workshops do VI Congresso Brasileiro de Informática na Educação (WCBIE 2017)*, 1:724. Minas Gerais, Brasil. <https://doi.org/10.5753/cbie.wcbie.2017.724>.
- Sánchez, J. L., F. Serradilla, E. Martínez, e J. Bobadilla. 2008. «Choice of metrics used in collaborative filtering and their impact on recommender systems». *2008 2nd IEEE International Conference on Digital Ecosystems and Technologies, IEEE-DEST 2008*, 432–36. <https://doi.org/10.1109/DEST.2008.4635147>.
- Schafer, B., Dan Frankowski, e Shilad Sen. 2014. «Collaborative Filtering Recommender Systems». Em *Lncs*, editado por P. Brusilovsky; A. W. Nejdl Kobsa; 4321:291–324. Springer-Verlag Berlin Heidelberg 2007. <https://dl.acm.org/citation.cfm?id=1768208>.

- Sheugh, L., e Sasan H. Alizadeh. 2015. «A note on pearson correlation coefficient as a metric of similarity in recommender system». Em *AI and Robotics, IRANOPEN 2015 - 5th Conference on Artificial Intelligence and Robotics*, 1–6. Qazvin, Irão: IEEE. <https://doi.org/10.1109/RIOS.2015.7270736>.
- Solima, L., Maria Rosaria Della Peruta, e Vincenzo Maggioni. 2016. «Managing adaptive orientation systems for museum visitors from an IoT perspective». *Business Process Management Journal* 2, no. 2: 285–304. <https://doi.org/10.1108/BPMJ-08-2015-0115>.
- Tsai, C. Y., e Bo Han Lai. 2014. «A Location-Item-Time sequential pattern mining algorithm for route recommendation». *Knowledge-Based Systems* 73: 97–110. <https://doi.org/10.1016/j.knosys.2014.09.012>.
- Wang, Y., Jiangzhou Deng, Jerry Gao, e Pu Zhang. 2017. «A hybrid user similarity model for collaborative filtering». *Information Sciences* 418–419: 102–18. <https://doi.org/https://doi.org/10.1016/j.ins.2017.08.008>.
- Ying, J. Jia-Ching, Eric Hsueh-Chan Lu, Wang-Chien Lee, Tz-Chiao Weng, e Vincent S. Tseng. 2010. «Mining user similarity from semantic trajectories». Em *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, 19–26. California: ACM Digital Library. <https://doi.org/10.1145/1867699.1867703>.
- Ying, J. Jia-Ching, Wang-Chien Lee, Tz-Chiao Weng, e Vincent S. Tseng. 2011. «Semantic trajectory mining for location prediction». Em *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 34–43. Illinois: ACM Digital Library. <https://doi.org/10.1145/2093973.2093980>.

Anexos

Pseudocódigo

$U = \{\text{utilizador}\}$

$POI = \{\text{Ponto de interesse}\}$

$u_i \in U, i = 1, \dots, \#U$

$POI_j \in POI, j = 1, \dots, \#POI$

Anexo 1 - Filtragem Colaborativa por trajetos

$T_semelhantes(u, \langle u_i, p_1, \dots, p_n \rangle, MedAval, S)$

$bestKAvT_u := 0$

$bestKU := NA$

$numU := 0$

$T_u :=$ trajetos realizados por u

$S := \{\}$

Para cada u_i com $i = 1, \dots, \#U$

$T_u_i :=$ trajetos realizados por u_i

$AvT_u_i := MedAval(T_u, T_u_i)$

Se $(numU < K)$

Então

$S := adiciona(S, \langle u_i, AvT_u_i \rangle)$

$bestKAvT_u := menorAvT_u(S)$

Se $(numU \geq K \text{ E } AvT_u_i > bestKAvT_u)$

Então

$S := adiciona(S, \langle u_i, AvT_u_i \rangle)$

$S := removemenorAvT_u(S)$

$bestKAvT_u := menorAvT_u(S)$

Retorna (S)

Anexo 2 - Filtragem colaborativa por pontos de interesse

$P_{\text{semelhantes}}(u, \langle u_i, p_1, \dots, p_n \rangle, \text{MedAval}, S)$

$\text{bestKAvT}_u := 0$

$\text{bestKU} := \text{NA}$

$\text{numU} := 0$

$P_u :=$ Pois visitados por u

$S := \{\}$

Para cada u_i com $i=1, \dots, \#U$

$P_{u_i} :=$ Pois visitados por u_i

$\text{AvP}_{u_i} := \text{MedAval}(P_u, P_{u_i})$

Se $(\text{numU} < K)$

Então

$S := \text{adiciona}(S, \langle u_i, \text{AvP}_{u_i} \rangle)$

$\text{bestKAvP}_u := \text{menorAvP}_u(S)$

Se $(\text{numU} \geq K \text{ E } \text{AvP}_{u_i} > \text{bestKAvP}_u)$

Então

$S := \text{adiciona}(S, \langle u_i, \text{AvP}_{u_i} \rangle)$

$S := \text{removemenorAvP}_u(S)$

$\text{bestKAvP}_u := \text{menorAvP}_u(S)$

Retorna (S)

Anexo 3 - Argumento de maximização da função

$\text{opt}(u, S, \text{maxTemp}, \text{maxPOIs}, \{\text{dist}(\text{POI1}, \text{POI2}), \{\text{tempo}(\text{POI1}, \text{POI2})\})$

$P_u :=$ Pois visitados por u

$P_S :=$ Pois visitados pelos utilizadores em S que não constam de P_u

$$\operatorname{argmax}_{\langle p_1, \dots, p_m \rangle} a \times m \times b \times \sum_{i=1}^{m-1} \operatorname{tempo}(POIs_{s_i}, POIs_{s_{i+1}}) + c \times \sum_{i=1}^{m-1} \operatorname{dist}(POIs_{s_i}, POIs_{s_{i+1}})$$

s.a:

$$m \leq \max POIs$$

$$\sum_{i=1}^{m-1} \operatorname{tempo}(POIs_{s_i}, POIs_{s_{i+1}}) \leq \max Tempo$$

$$\sum_{i=1}^{m-1} \operatorname{dist}(POIs_{s_i}, POIs_{s_{i+1}}) \leq \max Dist$$

$$0 \leq a, b, c \leq 1$$

$$a + b + c = 1$$

retorna $\langle p_1, \dots, p_m \rangle$