



Lucas Fernando Leite dos Santos

Revisão Sistemática de Literatura sobre Recomendação de Pontos de Interesse

São José dos Campos, SP

Lucas Fernando Leite dos Santos

Revisão Sistemática de Literatura sobre Recomendação de Pontos de Interesse

Trabalho de conclusão de curso apresentado ao
Instituto de Ciência e Tecnologia – UNIFESP,
como parte das atividades para obtenção do tí-
tulo de Bacharel em Ciência da Computação.

Universidade Federal de São Paulo – UNIFESP

Instituto de Ciência e Tecnologia

Bacharelado em Ciência da Computação

Orientador: Prof. Dr. Lilian Berton

São José dos Campos, SP

Dezembro de 2017

Lucas Fernando Leite dos Santos

Revisão Sistemática de Literatura sobre Recomendação de Pontos de Interesse

Trabalho de conclusão de curso apresentado ao
Instituto de Ciência e Tecnologia – UNIFESP,
como parte das atividades para obtenção do tí-
tulo de Bacharel em Ciência da Computação.

Trabalho aprovado em 14 de Dezembro de 2017:

Prof. Dr. Lilian Berton
Orientador

Professor
Convidado 1

Professor
Convidado 2

Professor
Convidado 3

São José dos Campos, SP
Dezembro de 2017

Este trabalho é dedicado aos meus pais Sergio e Cristina e aos meus amigos

Agradecimentos

A minha família, principalmente meus pais por sempre me apoiarem em todas minhas decisões e jornadas. Aos meus amigos, pelos momentos de distração e entretenimento. E aos meus professores e orientadora Professora Doutora Lilian Berton, pela paciência, ensinamentos e disposição.

Resumo

O Avanço da tecnologia dos celulares e a evolução e popularização das redes sociais fermentaram novos estudos que relacionam o conteúdo publicado por usuários e suas opiniões dentro destas redes para ofertar produtos de maneira personalizada e direta as pessoas. A integração dessas redes com sistemas de mapeamento global (GPS), permitem um novo passo na oferta de determinados produtos, capturando não somente opiniões mas também a dinâmica de localização geográfica referente ao momento que ela foi gerada, habilitando o direcionamento destas recomendações para pontos como restaurantes, parques, sorveterias e outros lugares classificados como pontos de interesse.

Este trabalho apresenta uma revisão sistemática da literatura sobre o tema de recomendação personalizada de pontos de interesse com o objetivo de determinar a diversidade de algoritmos para conhecimento e futura seleção para desenvolvimento de um sistema próprio visando classificar pontos de interesse de acordo com opiniões e gostos pessoais obtidos da rede social Twitter.

Palavras-chaves: Recomendação, Twitter, Aprendizado de máquina.

Abstract

The recent advance on smartphones technology and the evolution and popularization of social networks fermented new studies relationing the content published by users and their opinions inside those networks to offer products in a personalized and forward way to those people. The integration of those networks with global mapping systems (GPS), allow a new step on product offering, capturing not only opinions but also the geographical location dynamics refering to the moment which it was generated, enabling the targeting of those recommendations to places like restaurants, parks, ice cream shops and other places classified as points of interest.

This work presents a sistematic review of the literature about personalized recommendation of points of interests with the objective of determining the diversity of algorithms for acknowledgement and future selection to the development of an own system aiming to classify points of interest according with opinions and personal tastes acquired through social network Twitter.

Key-words: Recommendation, Twitter, Machine learning.

Lista de ilustrações

Figura 1 – Campos do conhecimento aplicados a Mineração de Dados	23
Figura 2 – Fluxograma de revisão sistemática	25
Figura 3 – Distribuição dos anos de publicação	29
Figura 4 – Distribuição de artigos por biblioteca digital	30
Figura 5 – Fatores analisados por cada artigo	31

Lista de tabelas

Tabela 1 – Algoritmos populares de aprendizado de máquina	22
Tabela 2 – Bibliotecas digitais utilizadas para pesquisa	26
Tabela 3 – Quantidade de artigos obtidos na busca e filtragem	26
Tabela 4 – Artigos obtidos após todas etapas de filtrações	27
Tabela 5 – Abordagem utilizada por cada publicação	32
Tabela 6 – Tabela de cronograma de atividades.	34

Lista de abreviaturas e siglas

LBSN	Location-Based Social Network
GSN	Geosocial network
API	Application programming interface

Sumário

1	Introdução	19
1.1	Contextualização	19
1.2	Motivação	19
1.3	Objetivos	20
1.3.1	Objetivo Geral	20
1.3.2	Objetivo Específico	20
2	Referencial Teórico	21
2.1	Geosocial Networking	21
2.1.1	Twitter	21
2.2	Aprendizado de Máquina	22
2.3	Mineração de Dados	23
3	Revisão Sistemática	25
3.1	Planejamento	25
3.2	Execução	26
3.3	Resultados e Análises	29
3.4	Conclusão e Considerações	32
4	Proposta para TCC-II	33
4.1	Atividades	33
4.2	Cronograma	34
	Referências	35

1 Introdução

Este capítulo tem como objetivo apresentar a contextualização, a motivação e os objetivos para a realização desta revisão sistemática da literatura, a respeito de recomendações personalizadas de pontos de interesse baseado em informações obtidas através da rede social Twitter.

1.1 Contextualização

As redes sociais são parte predominante na utilização da internet pela população geral, sua principal função é a interação e comunicação entre pessoas, seja por textos, imagens, vídeos ou qualquer outra forma de transmissão de mensagens. O Twitter é uma dessas redes, que tem como principal característica uma comunicação breve e direta através de uma quantidade limitada de 280 (previamente 140) caracteres e utilização de *hashtags* para classificação de temas. E sua utilização em grande parte é focada em opiniões e discussões diretas entre um usuário e seus seguidores, (KWAK et al., 2010).

Com o crescimento da utilização destas redes a internet se tornou um ambiente repleto de informações pessoais e opiniões que fermentam uma grande rede de dados disponíveis para ser utilizados para os mais diversos fins. Muitos estudos foram alimentados pela diversidade desses dados, sejam eles para uso comercial, entretenimento ou acadêmicos, os quais focam em diversos pontos desde comportamentais até estatísticos.

Um grande campo de pesquisa e desenvolvimento que vem crescendo em torno destes fatores é o uso de aprendizado de máquina para interpretação e classificação destes dados obtidos em postagens nestas redes. Todos estes estudos são feitos em cima da mineração destes dados de forma a poder reconhecer padrões para identificação de temas e objetivos diversos.

1.2 Motivação

Muitas discussões que populam o Twitter envolvem opiniões e recomendações de pontos de interesses (restaurantes, bares e locais para entretenimento de forma geral) focando o aspecto social destas redes, muitos estudos, (CHA et al., 2010) (KONSTAS; STATHOPOULOS; JOSE, 2009) (BAKSHY et al., 2011), mostram a influência envolvida nessas relações e discussões. Estas discussões geram naturalmente um ponto interessante para o campo de Data Mining, de forma a realizar extrações de informações relevantes. Dentre estas informações é possível obter citações, temas discutidos, coordenadas de localização da publicação e texto referente ao que a pessoa publicou possivelmente passando por interpretações textuais para

identificação de diversos fatores, como análise sentimental, identificação de nomes próprios entre outros.

No escopo das redes sociais algumas focam diretamente em recomendações e *reviews*, como o popular *TripAdvisor*, o que é muito atrativo para o público de forma geral por facilitar uma busca por entretenimento de forma mais direta e baseada na experiência de outros usuários, diferentemente de um sistema mais estático de recomendação onde o usuário busca por informações diretas do estabelecimento. Porém ambas abordagens não levam em consideração os interesses do usuário em particular, abrindo um espaço para inovação com utilização de ferramentas como a inteligência artificial para aplicação destas personalizações.

1.3 Objetivos

1.3.1 Objetivo Geral

Dentro deste contexto este trabalho buscará identificar e estudar formas de extrair dados relevantes da rede social Twitter que possam atuar como classificadores de tais pontos de interesse, focando em regiões e cidades através do uso de geolocalização. Posteriormente aplicar estas técnicas e desenvolver um sistema de classificação e recomendação personalizada para divulgação destes pontos, permitindo que as pessoas possam se identificar com determinados ambientes e locais de forma mais particular e expressiva.

1.3.2 Objetivo Específico

O principal objetivo da primeira etapa é realizar uma revisão sistemática de acordo com (MIAN et al., 2005), de forma a estudar o que é relevante atualmente, no escopo deste trabalho, e determinar as principais técnicas e ferramentas necessárias para o desenvolvimento de um sistema para recomendações de pontos de interesse.

A segunda etapa do trabalho consiste em desenvolver determinada ferramenta fundamentada nas referências determinadas como relevantes na primeira etapa e verificar sua eficácia de acordo com resultados esperados.

2 Referencial Teórico

Neste capítulo serão abordados conceitos e ferramentas relevantes ao tema deste trabalho de forma a facilitar e familiarizar-se com o escopo da pesquisa.

2.1 Geosocial Networking

As Redes Sociais são plataformas online que surgiram com a evolução da *web* com o intuito de gerar comunicações e gerenciar conexões entre pessoas através de perfis sociais (BOYD; ELLISON, 2007). A possibilidade de compartilhar opiniões e conteúdos em conjunto com a facilitação na comunicação, seja entre conhecidos na vida real ou não, fez com que estas redes tivessem rápida expansão de usuários que foi catalisada pela vinda dos *smartphones* e utilização de dados móveis (BLACKSTOCK; LEA; FRIDAY, 2011).

Dentre as redes sociais existentes temos diversas plataformas com os mais diversos usos, como o Facebook que agrega todo tipo de conteúdo e o Instagram e sua rede de fotografias. Com o aumento na diversidade destas redes e avanço das redes móveis e a tecnologia de GPS nos *smartphones* temos o surgimento das GSN (*Geosocial Networks*) que fazem uso da posição geográfica do usuário para que ele possa compartilhar acontecimentos ou opiniões divulgando sua localização atual em conjunto com o conteúdo gerado, como por exemplo quando o usuário divulga uma foto em seu perfil no Instagram e compartilha a localização na qual ele acabou de fotografar.

2.1.1 Twitter

O Twitter é uma das rede social que pertence ao grupo das GSN, famosa por criar interações rápidas entre usuários e seus seguidores através de apenas 280 caracteres, previamente 140. Dentro desta rede a conexão entre pessoas é formada por seguidores, onde usuários recebem em sua *timeline* o que todas as pessoas na qual eles seguem postam, em forma das mensagens chamadas *tweets*. Este vínculo é gerado quando um usuário valoriza a opinião de determinada pessoa e a julga relevante para que apareça em seu *feed* de *tweets*. Os *tweets* por sua vez possuem os assuntos mais variados possíveis, e além dos 280 caracteres ele pode conter *hashtags*, que são etiquetas para os *tweets* geralmente composta do caracter # seguido de um termo que ajuda na classificação e facilita na filtragem dos *tweets*.

Estas características tornam o Twitter uma interessante ferramenta de estudo e desenvolvimento. Suas *hashtags*, mensagens curtas, sua relevância e sua API (TWITTER..., 2017b) de fácil utilização, tornam esta rede a mais adequada para o desenvolvimento deste trabalho sobre sistema de recomendação.

2.2 Aprendizado de Máquina

Como subcampo da Inteligência Artificial que trata do reconhecimento de padrões e aprendizado computacional, o aprendizado de máquinas, como definido em (CARBONELL; MICHALSKI; MITCHELL, 1983), tem três focos principais para o desenvolvimento destes itens, são eles:

- Desenvolvimento e análise de sistemas de aprendizado para melhoria de performance em determinadas tarefas;
- Investigação e simulação computacional do processo de aprendizado humano;
- Exploração dos possíveis métodos de aprendizagem e algoritmos independentemente da aplicação.

De forma mais prática o aprendizado de máquinas possui algumas categorizações. As principais quanto a forma de aprendizado ele pode ser supervisionado, no qual exemplos de entradas e saídas são fornecidos ao algoritmo para que ele aprenda uma regra geral, ou não supervisionado, onde o algoritmo fica responsável em atingir seu objetivo sem que as entradas sejam fornecidas para treino. Também aplica-se os dois princípios anteriores nos algoritmos semi-supervisionados, para aproveitar dados que possam ser utilizados como exemplos para o treinamento do algoritmo. Ou então aprendizado por reforço, onde o sistema é retroalimentado a com positivas ou negativas a partir de um resultado obtido por ele mesmo.

Por outro lado também existem categorizações quanto ao resultado produzido pelos algoritmos. Para o problema de recomendação é desejado um resultado com um conjunto de dados categorizado e definido, os métodos que nos apresentam estes resultados são os definidos como agrupamento e classificação. O agrupamento (*clustering*) é o problema de se dividir um conjunto de dados em diferentes grupos através de reconhecimento de padrões e características deste conjunto (JAIN; MURTY; FLYNN, 1999). Já a classificação é a abordagem supervisionada do mesmo problema (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007), porém um conjunto de dados já classificado como pertencente a um determinado grupo, é utilizado como treinamento para que o algoritmo compare as novas entradas e as classifique. Na Tabela 1 temos alguns dos algoritmos mais populares dos métodos descritos.

Tabela 1 – Algoritmos populares de aprendizado de máquina

Categorização	Algoritmos
Agrupamento (supervisionado)	k-means, BIRCH e Maximização de Expectativas
Classificação (não supervisionado)	Árvore de Decisão, Rede Neural e kNN

2.3 Mineração de Dados

Devido a necessidade de uso de uma grande quantidade de informações para criação, aperfeiçoamento e análise de um sistema de recomendação, um alto fluxo de dados deve ser extraído de sua fonte e analisado de forma a obter-se o que é relevante ou não para este trabalho. A mineração de dados é o campo que estuda técnicas e metodologias para esta tarefa.

Para que a mineração de dados seja realizada, diversos campos matemáticos e computacionais são combinados para que o resultado obtido seja satisfatório e significativo. Como estatísticas, aprendizado de máquinas, sistemas de bancos de dados e outros, conforme visto na Figura 1.

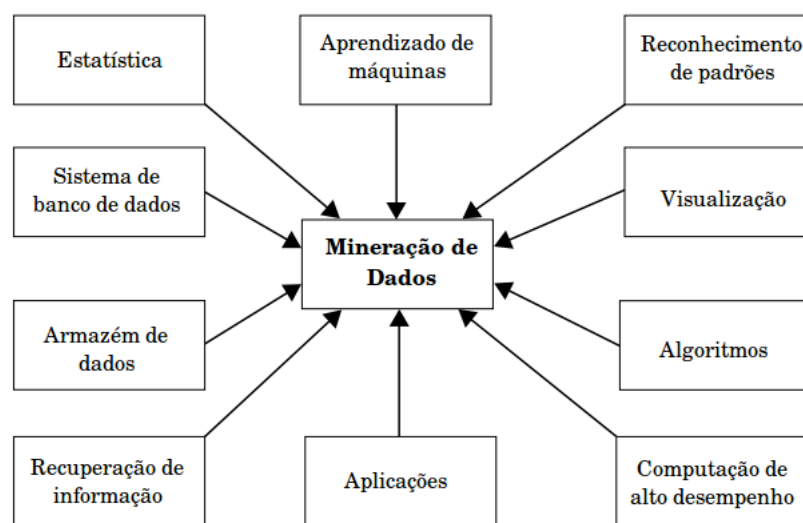


Figura 1 – Campos do conhecimento aplicados a Mineração de Dados. Traduzido de (HAN; KAMBER; PEI, 2006)

Segundo (HAND; MANNILA; SMYTH, 2001), esta mineração é realizada através de 5 etapas. A primeira é a exploração dos dados, que consiste em entender o conjunto de dados. A segunda é a criação de um modelo descritivo com o intuito de segmentar e agrupar os dados. Em seguida, na terceira etapa, um modelo de predição é criado para classificação e regressão das variáveis categoricamente e quantitativamente. Já a quarta etapa consiste em determinações de padrões e regras de acordo com a necessidade da aplicação dos dados. E por fim, na quinta etapa, temos a recuperação dos dados que serão utilizados, analisados ou estudados.

3 Revisão Sistemática

Para melhor compreensão do estado da arte neste capítulo será realizada uma revisão sistemática que servirá como base para o desenvolvimento do sistema.

A revisão sistemática é uma metodologia de revisão da literatura de determinado assunto iniciado na medicina com o objetivo dos médicos se manterem atualizados em suas especializações que rapidamente se espalhou para diversos outros campos acadêmicos, inclusive na computação.

Neste trabalho são seguidas as diretrizes propostas por (MIAN et al., 2005) e (PETERSEN et al., 2008), consistindo em etapas sequenciais como visto na Figura 2. A primeira etapa consiste em planejar, definir os objetivos e métodos de pesquisa. A segunda é a execução, onde os artigos são selecionados e filtrados de acordo com os critérios estabelecidos no planejamento. E por fim é feita a análise dos resultados onde os dados obtidos são sintetizados e estudados.

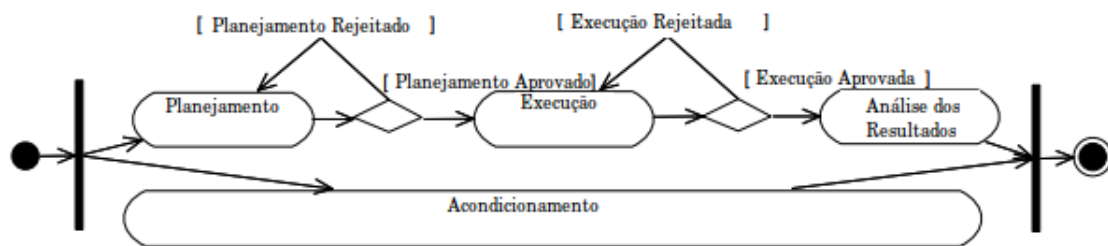


Figura 2 – Fluxograma de revisão sistemática. Traduzido de (MIAN et al., 2005)

3.1 Planejamento

O objetivo desta revisão é encontrar as mais eficazes ferramentas e métodos mais populares utilizados na recomendação de pontos de interesse através de dados minerados da rede social Twitter para uso como referência no desenvolvimento.

Com este objetivo definido temos um conjunto de palavras-chave que formam a *string* utilizada para a busca nas bibliotecas digitais, sendo elas: POI, Twitter e *recommendation*. Estes termos foram definidos em inglês devido a maior variedade de publicações disponíveis. Desta forma formamos a *string* correspondente ao padrão de busca das bibliotecas digitais e utilizando operadores lógicos como:

("POI"OR "place"or "personalized") AND "Twitter"AND "Recommendation"

Dado o objetivo prático desta revisão, foram adotadas apenas três bibliotecas digitais para a pesquisa dos artigos acadêmicos listadas na Tabela 2.

Tabela 2 – Bibliotecas digitais utilizadas para pesquisa

Biblioteca Digital	Site
ACM Digital Library	https://dl.acm.org/
IEEE Xplore	https://ieeexplore.ieee.org/
ScienceDirect	https://www.sciencedirect.com/

Para que o processo seja completo também foram definidos alguns critérios para filtração dos artigos na busca. Somente artigos em inglês foram selecionados, assim como justificado na definição da string de busca. Também foram buscados artigos com data de publicação a partir de 2013 para que as tecnologias utilizadas sejam as representativas após o início da popularização das redes sociais.

3.2 Execução

As buscas foram realizadas no dia 02 de Outubro de 2017. Aplicando a metodologia definida no planejamento de filtragem inicial por ano foram obtidos uma quantidade total de 2436 de acordo com a Tabela 3, porém, um grande número de artigos adquiridos pela busca não possuem relação direta com o assunto desejado, por isso algumas etapas de filtragem destes artigos foram realizadas:

1. Leitura dos títulos para identificar artigos com assuntos muito discrepantes do desejado;
2. Leitura dos resumos para entendimento do escopo do artigo e validação de relevância. Aqui nesta etapa foram removidos artigos que não se tratavam da rede social Twitter ou faziam recomendação de sistemas não relacionados a pontos de interesse;
3. Leitura dos detalhes de cada artigo, de forma a determinar se havia relação direta com o objetivo deste trabalho e diferenciar detalhes que não puderam ser compreendidos durante a leitura do resumo.

Tabela 3 – Quantidade de artigos obtidos na busca e filtragem

Biblioteca	Total	1ª Filtragem	2ª Filtragem	3ª Filtragem
ACM	135	48	7	4
IEEE Xplore	138	79	18	9
ScienceDirect	2163	56	15	5

Após todas as etapas de filtragem cada artigo foi listado como visto na Tabela 4 onde consta o título, autores e ano de publicação.

Tabela 4 – Artigos obtidos após todas etapas de filtragens

ID	Título	Autores	Ano
A1	Social Recommender System: A Recommender System Based on Tweets for Points of Interest (GARCÍA et al., 2017)	Cristian Gonzalez García and Daniel Meana-Llorian and Vicente Garcaz and Edward Rolando Núñez-Valdez	2017
A2	Can We Predict Eat-out Preference of a Person from Tweets? (RAHMAN et al., 2016)	Md. Mahabur Rahman and Md Tak-sir Hasan Majumder and Md Saddam Hossain Mukta and Mohammed Eunus Ali and Jalal Mahmud	2016
A3	Assessing the Contribution of Twitter’s Textual Information to Graph-based Recommendation (PRITSKER; KUFLIK; MINKOV, 2017)	Evgenia Wasserman Pritsker and Tsvi Kuflik and Einat Minkov	2017
A4	Annotating Points of Interest with Geo-tagged Tweets (ZHAO; CONG; SUN, 2016)	Kaiqi Zhao and Gao Cong and Aixin Sun	2016
A5	A General Geographical Probabilistic Factor Model for Point of Interest Recommendation (LIU et al., 2015)	B. Liu; H. Xiong; S. Papadimitriou; Y. Fu; Z. Yao	2015
A6	Social tagging recommendation system for smart city environments (ALHAMID, 2016)	M. F. Alhamid	2016
A7	Context-aware point of interest recommendation using tensor factorization (MAROULIS; BOUTSIS; KALOGERAKI, 2016)	S. Maroulis; I. Boutsis; V. Kalogeraki	2016
A8	Point-of-Interest Recommendations via a Supervised Random Walk Algorithm (XU; FU; GU, 2016)	G. Xu; B. Fu; Y. Gu	2016

A9	Adaptive location recommendation algorithm based on location-based social networks (LIN et al., 2015)	K. Lin; J. Wang; Z. Zhang; Y. Chen; Z. Xu	2015
A10	Topic-Sensitive Location Recommendation with Spatial Awareness (GUO; HUANG; THENG, 2015)	Q. Guo; Y. Huang; Y. L. Theng	2015
A11	Point-of-interest recommendation in location-based social networks with personalized geo-social influence (LIWEI; YUTAO; YANBO, 2015)	L. Huang; Y. Ma; Y. Liu	2015
A12	SAR: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews (ZHAO et al., 2015)	K. Zhao; G. Cong; Q. Yuan; K. Q. Zhu	2015
A13	Semantic-Based Location Recommendation With Multimodal Venue Semantics (WANG et al., 2015)	X. Wang; Y. L. Zhao; L. Nie; Y. Gao; W. Nie; Z. J. Zha; T. S. Chua	2015
A14	GeoSRS: A hybrid social recommender system for geo-located data (CAPDEVILA; ARIAS; ARRATIA, 2016)	Capdevila, Joan; Arias, Marta and Arratia, Argimiro	2016
A15	Context-aware probabilistic matrix factorization modeling for point-of-interest recommendation (REN et al., 2017)	Ren, Xingyi; Song, Meina; E, Haihong and Song, Junde	2017
A16	CTF-ARA: An adaptive method for POI recommendation based on check-in and temporal features (SI; ZHANG; LIU, 2017)	Si, Yali; Zhang, Fuzhi and Liu, Wenyan	2017

A17	Textual-geographical-social aware point-of-interest recommendation (XINGYI et al., 2016b)	Xingyi, Ren; Meina, Song; Haihong, E and Junde, Song	2016
A18	Joint model of user check-in activities for point-of-interest recommendation (XINGYI et al., 2016a)	Xingyi, Ren; Meina, Song; Haihong, E and Junde, Song	2016

3.3 Resultados e Análises

Analisando a distribuição dos anos de publicação no gráfico da Figura 3 é notável que o assunto tem um apelo recente, mesmo com as buscas sendo realizadas para artigos a partir de 2013 apenas artigos publicados entre 2015 e 2017 passaram pela última etapa de filtragem. Esse recente interesse em desenvolvimento em sistemas de recomendações é fomentado pela evolução da tecnologia dos *smartphones*, GPS e popularização do uso das redes sociais.

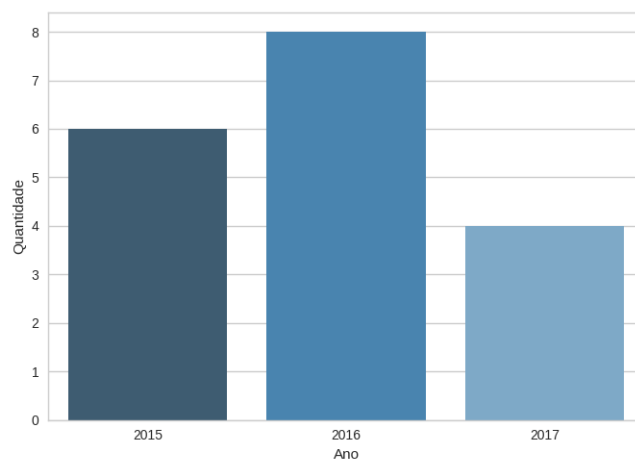


Figura 3 – Distribuição dos anos de publicação

No gráfico da Figura 4 pode notar-se que a biblioteca IEEE Xplore resultou em mais artigos aproveitados apesar da biblioteca Science Direct possuir uma quantidade muito maior de artigos para a string de busca aplicada. Parte desta diferença se deve ao fato de que grande parte dos artigos da Science Direct são de áreas não relacionadas diretamente com a computação, principalmente medicina, que logo na primeira etapa da filtragem foram descartados como visto na Tabela 3.

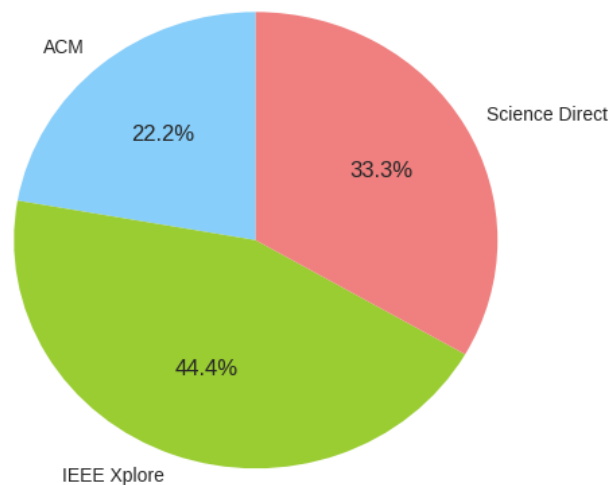


Figura 4 – Distribuição de artigos por biblioteca digital

Dentre os dados minerados cada método aplica análises e classificações de acordo com o que ele determina mais relevante e determinante. A Figura 5 mostra a distribuição de artigos dentro de cada fator foi levado em consideração.

A geolocalização é o fator mais predominante, pois pode ser obtido através da localização direta de uma geotag, que fornece as coordenadas geográficas, ou através de um check-in feito pelo usuário no qual proporciona a identificação do POI no qual ele se encontra de forma direta. O fator geográfico além de permitir a identificação do POI também é utilizado para determinar a frequência na qual o usuário frequenta determinado POI, formando assim um indício de qualidade e preferência para o sistema de recomendação.

A análise textual também é extremamente relevante pois diversos métodos, 12 dentre os artigos selecionados, utilizam destas análises também para determinar localizações, caso não exista a geotag ou o check-in no Tweet. Porém a principal aplicação da análise textual está na classificação dos POI e em suas avaliações, estas classificações visam determinar características específicas sobre o POI, seja ele um restaurante ou uma sorveteria por exemplo, e as avaliações são feitas através de análises de sentimento e determinadas palavras que podem indicar a qualidade do POI.

Por fim o fator de menos utilizado pelos sistemas descritos no artigo é o fator social, simplificada ele é considerado de duas formas. A primeira é similaridade entre diferentes usuários que podem possuir os mesmos interesses, a segunda consideração é o interesse entre as pessoas conectadas nas redes, no caso do Twitter entre o usuário e as pessoas que ele segue.

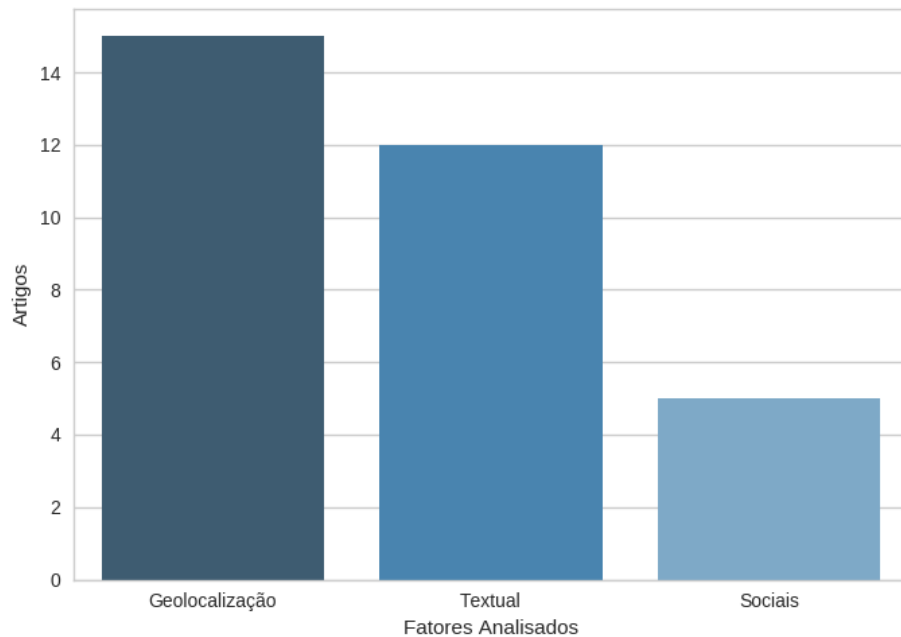


Figura 5 – Fatores analisados por cada artigo

Com relação a abordagem técnica utilizadas nos artigos, alguns se dividem em várias etapas e fazem utilização de diversas técnicas diferentes para classificação, agrupamento ou regressão. Na Tabela 5 temos a lista com os principais algoritmos utilizados por cada publicação. Esta tabela representa somente a principal abordagem que cada artigo propôs, como mencionado, muitas delas são acompanhadas de diversas outras etapas importantes, como por exemplo análise de sentimentos e similaridades, ou então constituem-se de variações próprias de dadas metodologias, formando um sistema complexo que leva em consideração os diferentes fatores apresentados anteriormente.

Dentro desta listagem verifica-se a variedade de possibilidades para desenvolver o problema de recomendação através do aprendizado de máquinas, a eficácia de cada um desses métodos variam de acordo com sua aplicação e personalização desejada. A Tabela 5 também mostra quais fatores cada uma das abordagens apresentadas nos artigos levaram em consideração na construção do sistema.

Tabela 5 – Abordagem utilizada por cada publicação

ID	Abordagem	Fator(es) Considerado(s)
A1	Classificação (algoritmo não especificado)	Texto
A2	Árvore de Regressão	Texto e geolocalização
A3	Random Walk Personalized PageRank	Texto e social
A4	Naive Bayesian Classifier	Texto e geolocalização
A5	Probabilistic Factor Model	Geolocalização
A6	Classificação (algoritmo não especificado)	Texto e geolocalização
A7	Non-negative Tensor Factorization	Geolocalização
A8	Supervised Random Walk	Texto e social
A9	Naive Bayesian Classifier	Geolocalização
A10	Probabilistic Model PageRank	Texto e geolocalização
A11	Probabilistic Factor Graph Model	Geolocalização
A12	Probabilistic Model	Texto e geolocalização
A13	k-NN	Texto e geolocalização
A14	k-NN	Texto e geolocalização
A15	Matrix Factorization	Texto, geolocalização e social
A16	k-Means	Geolocalização
A17	Matrix Factorization	Texto, geolocalização e social
A18	k-Means	Geolocalização e social

3.4 Conclusão e Considerações

Apesar de existirem muitos estudos propondo novos métodos e tecnologias envolvendo sistemas de recomendação, a aplicação de um sistema para recomendação de pontos de interesse implica também na utilização do fator geográfico, diferente de recomendações de filmes ou música que podem levar somente fatores sociais e textuais em consideração. Dada esta limitação, o número de artigos e publicações encontrados neste campo não foram muito volumosos, porém, este campo possui grande apelo comercial, como pode ser notado na expansão de redes sociais como Foursquare e Yelp que fazem uso de recomendações para seus usuários.

4 Proposta para TCC-II

A segunda parte deste trabalho consiste no desenvolvimento de um sistema de recomendação para pontos de interesse utilizando como referência os artigos revisados durante esta primeira etapa. Estes artigos fornecem uma base teórica do que é atual e efetivo na aplicação a ser desenvolvida.

Para que a ferramenta seja desenvolvida as etapas listadas deverão ser cumpridas:

- Estudo detalhado da bibliografia para escolha de métodos e algoritmos;
- Estudo de ferramentas de mineração de dados;
- Desenvolvimento de ferramenta para mineração de Tweets;
- Desenvolvimento do algoritmo para classificação dos POI;
- Desenvolvimento do sistema de recomendação.
- Estudo de eficiência e eficácia do sistema;

4.1 Atividades

1. Seleção de dados e montagem do conjunto de treinamento e teste

Será utilizada a API do Twitter ([TWITTER...](#), 2017a) através da biblioteca Tweepy ([TWEETPY](#), 2017) para a coleta de *tweets* com geolocalização de São José dos Campos.

2. Pré-processamento dos dados

Os dados coletados serão pré-processados com a biblioteca NLTK ([NATURAL...](#), 2017) disponível para *Python*, removendo *stopwords*, sinais de pontuação e lematizando as palavras. Alguns *tweets* serão agrupados e rotulados manualmente de acordo com a geolocalização (lazer, alimentação, compras, esportes, etc.) e com classificações etárias, horário, positivo/negativo e outras.

3. Análise de algoritmos para classificação dos dados

Algoritmos como o *Word2vec* serão aplicados para gerar vetores de similaridade entre *tweets* e para aplicá-los em algoritmos de classificação. Os algoritmos serão treinados com os dados gerados manualmente na etapa anterior e aplicados nos dados de teste. Diversos algoritmos de aprendizado de máquinas disponíveis para *Python* serão testados e avaliados.

4. Teste da ferramenta para recomendações

Após a finalização dos algoritmos será desenvolvido um buscador simples, onde o usuário entra com a *query* que deseja buscar recomendação, por exemplo "restaurante", "sorveteria", etc., a ferramenta então retornará recomendações de lugares de acordo com as classificações geradas.

5. Redação da monografia

Por fim a monografia será redigida, levando em consideração a revisão bibliográfica feita nesta primeira parte do TCC e a ferramenta que será desenvolvida na segunda parte do TCC.

6. Defesa do TCC

A defesa do TCC está prevista para julho de 2018 com banca a ser definida.

4.2 Cronograma

As atividades do TCC 2 serão realizadas de acordo com o cronograma da tabela 6.

Tabela 6 – Tabela de cronograma de atividades.

Atividade	Fevereiro	Março	Abril	Maió	Junho	Julho
1	X					
2	X	X				
3		X	X	X		
4				X	X	
5					X	X
6						X

Referências

- ALHAMID, M. F. Social tagging recommendation system for smart city environments. In: IEEE. *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. [S.l.], 2016. p. 1–6. Citado na página 27.
- BAKSHY, E. et al. Everyone’s an influencer: quantifying influence on twitter. In: ACM. *Proceedings of the fourth ACM international conference on Web search and data mining*. [S.l.], 2011. p. 65–74. Citado na página 19.
- BLACKSTOCK, M.; LEA, R.; FRIDAY, A. Uniting online social networks with places and things. In: *Proceedings of the Second International Workshop on Web of Things*. New York, NY, USA: ACM, 2011. (WoT ’11), p. 5:1–5:6. ISBN 978-1-4503-0624-9. Disponível em: <<http://doi.acm.org/10.1145/1993966.1993974>>. Citado na página 21.
- BOYD, d. m.; ELLISON, N. B. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, Blackwell Publishing Inc, v. 13, n. 1, p. 210–230, 2007. ISSN 1083-6101. Disponível em: <<http://dx.doi.org/10.1111/j.1083-6101-2007.00393.x>>. Citado na página 21.
- CAPDEVILA, J.; ARIAS, M.; ARRATIA, A. Geosrs: A hybrid social recommender system for geolocated data. *Information Systems*, Elsevier, v. 57, p. 111–128, 2016. Citado na página 28.
- CARBONELL, J. G.; MICHALSKI, R. S.; MITCHELL, T. M. An overview of machine learning. In: *Machine learning*. [S.l.]: Springer, 1983. p. 3–23. Citado na página 22.
- CHA, M. et al. Measuring user influence in twitter: The million follower fallacy. In: *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*. [S.l.: s.n.], 2010. Citado na página 19.
- GARCÍA, C. G. et al. Social recommender system: A recommender system based on tweets for points of interest. In: ACM. *Proceedings of the 4th Multidisciplinary International Social Networks Conference on ZZZ*. [S.l.], 2017. p. 28. Citado na página 27.
- GUO, Q.; HUANG, Y.; THENG, Y.-L. Topic-sensitive location recommendation with spatial awareness. In: IEEE. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2015 IEEE/WIC/ACM International Conference on*. [S.l.], 2015. v. 1, p. 237–243. Citado na página 28.
- HAN, J.; KAMBER, M.; PEI, J. *Data mining: Concept and techniques Ed ke-2*. [S.l.]: San Francisco: Morgan kaufman Publisher, 2006. Citado na página 23.
- HAND, D. J.; MANNILA, H.; SMYTH, P. *Principles of data mining*. [S.l.]: MIT press, 2001. Citado na página 23.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. *ACM computing surveys (CSUR)*, Acm, v. 31, n. 3, p. 264–323, 1999. Citado na página 22.

KONSTAS, I.; STATHOPOULOS, V.; JOSE, J. M. On social networks and collaborative recommendation. In: ACM. *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. [S.l.], 2009. p. 195–202. Citado na página 19.

KOTSIANTIS, S. B.; ZAHARAKIS, I.; PINTELAS, P. *Supervised machine learning: A review of classification techniques*. 2007. Citado na página 22.

KWAK, H. et al. What is twitter, a social network or a news media? In: *Proceedings of the 19th International Conference on World Wide Web*. New York, NY, USA: ACM, 2010. (WWW '10), p. 591–600. ISBN 978-1-60558-799-8. Citado na página 19.

LIN, K. et al. Adaptive location recommendation algorithm based on location-based social networks. In: IEEE. *Computer Science & Education (ICCSE), 2015 10th International Conference on*. [S.l.], 2015. p. 137–142. Citado na página 28.

LIU, B. et al. A general geographical probabilistic factor model for point of interest recommendation. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 27, n. 5, p. 1167–1179, 2015. Citado na página 27.

LIWEI, H.; YUTAO, M.; YANBO, L. Point-of-interest recommendation in location-based social networks with personalized geo-social influence. *China Communications*, IEEE, v. 12, n. 12, p. 21–31, 2015. Citado na página 28.

MAROULIS, S.; BOUTSIS, I.; KALOGERAKI, V. Context-aware point of interest recommendation using tensor factorization. In: IEEE. *Big Data (Big Data), 2016 IEEE International Conference on*. [S.l.], 2016. p. 963–968. Citado na página 27.

MIAN, P. et al. A systematic review process for software engineering. In: *ESELAW'05: 2nd Experimental Software Engineering Latin American Workshop*. [S.l.: s.n.], 2005. Citado 2 vezes nas páginas 20 e 25.

NATURAL Language Toolkit. 2017. Disponível em: <<http://www.nltk.org/>>. Citado na página 33.

PETERSEN, K. et al. Systematic mapping studies in software engineering. In: *EASE*. [S.l.: s.n.], 2008. v. 8, p. 68–77. Citado na página 25.

PRITSKER, E. W.; KUFLIK, T.; MINKOV, E. Assessing the contribution of twitter's textual information to graph-based recommendation. In: *Proceedings of the 22Nd International Conference on Intelligent User Interfaces*. New York, NY, USA: ACM, 2017. (IUI '17), p. 511–516. ISBN 978-1-4503-4348-0. Disponível em: <<http://doi.acm.org/10.1145/3025171-3025218>>. Citado na página 27.

RAHMAN, M. M. et al. Can we predict eat-out preference of a person from tweets? In: *Proceedings of the 8th ACM Conference on Web Science*. New York, NY, USA: ACM, 2016. (WebSci '16), p. 350–351. ISBN 978-1-4503-4208-7. Disponível em: <<http://doi.acm.org/10.1145/2908131.2908199>>. Citado na página 27.

REN, X. et al. Context-aware probabilistic matrix factorization modeling for point-of-interest recommendation. *Neurocomputing*, Elsevier, v. 241, p. 38–55, 2017. Citado na página 28.

- SI, Y.; ZHANG, F.; LIU, W. Ctf-ara: An adaptive method for poi recommendation based on check-in and temporal features. *Knowledge-Based Systems*, Elsevier, v. 128, p. 59–70, 2017. Citado na página 28.
- TWEETPY. 2017. Disponível em: <<http://tweepy.readthedocs.io/en/v3.5.0/index.html>>. Citado na página 33.
- TWITTER API. 2017. Disponível em: <<https://developer.twitter.com/>>. Citado na página 33.
- TWITTER, Inc. 2017. Disponível em: <<https://about.twitter.com/company>>. Citado na página 21.
- WANG, X. et al. Semantic-based location recommendation with multimodal venue semantics. *IEEE Transactions on Multimedia*, IEEE, v. 17, n. 3, p. 409–419, 2015. Citado na página 28.
- XINGYI, R. et al. Joint model of user check-in activities for point-of-interest recommendation. *The Journal of China Universities of Posts and Telecommunications*, Elsevier, v. 23, n. 4, p. 25–36, 2016. Citado na página 29.
- XINGYI, R. et al. Textual-geographical-social aware point-of-interest recommendation. *The Journal of China Universities of Posts and Telecommunications*, Elsevier, v. 23, n. 6, p. 24–67, 2016. Citado na página 29.
- XU, G.; FU, B.; GU, Y. Point-of-interest recommendations via a supervised random walk algorithm. *IEEE Intelligent Systems*, IEEE, v. 31, n. 1, p. 15–23, 2016. Citado na página 27.
- ZHAO, K.; CONG, G.; SUN, A. Annotating points of interest with geo-tagged tweets. In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2016. (CIKM '16), p. 417–426. ISBN 978-1-4503-4073-1. Disponível em: <<http://doi.acm.org/10.1145/2983323.2983850>>. Citado na página 27.
- ZHAO, K. et al. Sar: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews. In: *IEEE. Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. [S.l.], 2015. p. 675–686. Citado na página 28.