



Universidade Federal de Pernambuco
Centro de Informática

Curso de Ciência da Computação

**Sistema de recomendação de localização para Redes
Sociais Baseadas em Localização**

Victor Augusto Pereira Porciúncula

Trabalho de Graduação

Recife,
Junho, 2018

Universidade Federal de Pernambuco
Centro de Informática

Victor Augusto Pereira Porciúncula

**Sistema de recomendação de localização para Redes
Sociais Baseadas em Localização**

Trabalho apresentado ao Centro de Informática da Universidade federal de pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Professor Orientador: Ricardo Bastos Cavalcante Prudêncio

Recife
Junho, 2018

Agradecimentos

Gostaria de agradecer a minha mãe por sempre ter acreditado em mim e ter sempre tido minha educação como prioridade, tudo que conquistei até hoje foi com o auxílio dela desde o começo da minha formação até a graduação.

Gostaria de agradecer a minha avó por sempre estar do meu lado em momentos difíceis e ter me dado o devido suporte nestes, além de ser um exemplo de pessoa gentil e bondosa na minha vida.

Obrigado aos meus colegas e amigos que fizeram parte da minha trajetória acadêmica e que de alguma forma me trouxeram aprendizado, seja dentro da faculdade ou estágio, ou fora.

Obrigado a todos os professores que fizeram parte da minha caminhada dentro da universidade.

Agradeço ao professor Ricardo Bastos Cavalcante Prudêncio pela orientação neste trabalho de graduação.

A educação é a arma mais poderosa que você pode usar para mudar o mundo

Nelson Mandela

Resumo

A utilização de redes sociais já é algo que é inerente a nossa sociedade; com o desenvolvimento e fácil acessibilidade da rede, a adição de localização nas redes sociais já é muito comum. Além das redes sociais baseadas em localização, redes sociais, como o Instagram, utiliza localização para por exemplo, apenas marcar conteúdos (fotos e vídeos).

Este trabalho visa estudar e categorizar os tipos de sistemas de recomendação de localização para redes sociais baseadas em localização. A partir deste estudo foi selecionado um artigo a servir como guia para a implementação, que no caso foi o *User Preferences-Based and Time-Sensitive Location Recommendation Using Check-In Data*, este propõe um modelo de fusão que leva em conta quatro modelos com as seguintes características: popularidade da localização, preferência do usuário, o momento de visita do usuário e os interesses dos amigos.

A partir da implementação de [3], foi feito um experimento com uma parte do dataset da antiga rede social Brightkite [14], essa subdivisão contém 500 usuários e 255.414 check-ins, e destes 500 foram selecionados aleatoriamente 100 usuários. Os resultados mostraram que obteve-se um melhor desempenho utilizando apenas o modelo de preferência do usuário, este desempenho foi abaixo do experimento realizado em [3], entretanto, ainda superior aos outros algoritmos comparados em [3].

Sumário

| | |
|--|-----------|
| 1. Introdução | 9 |
| 1.1. Motivação | 9 |
| 1.2. Objetivo | 10 |
| 1.3. Metodologia | 10 |
| 1.4. Estrutura do Trabalho | 10 |
| 2. Revisão de Literatura | 11 |
| 2.1. Sobre redes sociais baseadas em localização | 11 |
| 2.2. Contexto e compreensão da adição de localização às redes sociais | 11 |
| 2.3. Propriedades das redes sociais baseadas em localização | 13 |
| 2.4. Categorizações dos sistemas de recomendação | 14 |
| 2.4.1. Categorização dos sistemas de recomendação por objetivo | 14 |
| 2.4.2. Categorização dos sistemas de recomendação por metodologia | 15 |
| 2.4.3. Categorização dos sistemas de recomendação por fonte de dados | 17 |
| 3. Método Proposto | 18 |
| 3.1. Recomendação sensível ao tempo e baseada em preferência de usuário. | 18 |
| 3.2. Fatores considerados no sistema de recomendação. | 19 |
| 3.3. Modelos de recomendação de localização | 20 |
| 3.3.1. Características temporais e popularidade do local | 20 |
| 3.3.2. Preferência do usuário | 21 |
| 3.3.3. Influência Social dos amigos | 22 |
| 3.3.4. Modelo de Fusão | 23 |
| 4. Experimentos | 24 |
| 4.1. Dados | 24 |
| 4.2. Avaliação | 24 |
| 4.3. Resultados | 26 |
| 5. Conclusão | 28 |
| Referências | 29 |

Lista de Figuras

1. Funções Distributivas Acumulativas Complementares para o número de check-ins nos lugares (esquerda) e o número de check-ins por usuários (direita) [5].
2. Plot empilhado das 10 categorias mais populares sobre os dias úteis e os fins de semana. A popularidade decresce de cima para baixo [5].
3. Propriedades nas redes sociais baseadas em localização [2].
4. Visão geral sobre as categorias de sistemas de recomendação para redes sociais baseadas em localização [2].
5. Estrutura de Rede Bayesiana para inferir preferência do usuário [9].
6. O conceito básico do modelo HITS [10].
7. Distribuição de check-ins da (a) localização 153.505, (b) 14.710 [3].
8. O número de datas que o (a) usuário 1251 visitou e (b) o usuário 2 visitou [3].
9. Precision utilizando $\alpha = 0$ e $\beta = 1$.
10. Recall utilizando $\alpha = 0$ e $\beta = 1$.
11. F-measure utilizando $\alpha = 0$ e $\beta = 1$.

Lista de Tabelas

1. Resultado em precision/recall para cada par de possibilidades, com $N=1$.
2. Resultado em f-measure para cada par de possibilidades, com $N=1$.
3. Resultado em precision/recall para cada par de possibilidades α e $\beta \in \{0, 0.33, 0.5, 1\}$.
4. Resultado em f-measure para cada par de possibilidades α e $\beta \in \{0, 0.33, 0.5, 1\}$.

1. Introdução

1.1. Motivação

Redes sociais como Twitter, Instagram e Facebook vêm se tornando as aplicações na Internet com maior popularidade, tanto em números de usuários como em engajamento. O acúmulo de dados com o passar dos anos por estas redes sociais permitiu o surgimento de vários tipos de sistemas de recomendação, tanto de amigos, como de mídia (fotos, músicas, vídeos, etc).

Com o progresso da facilidade de adquirir informações de locais e da comunicação wireless, surgiram aplicações como Foursquare, Gowalla e Whrrl. Redes sociais combinadas com dados geográficos evoluíram para redes sociais geográficas [1]. Estas aplicações permitem ao usuário que este compartilhe quando e aonde foi, assim como algum conteúdo relacionado a experiência que teve no local (ex. fotos). Por exemplo, um usuário pode compartilhar em forma de comentário e imagem com seus amigos como foi sua experiência ao visitar um bar.

Além da experiência nos aplicativos que tem como objetivo principal o compartilhamento de geolocalização, tem-se essa possibilidade em aplicações sem este propósito como central, como o Instagram, onde você pode marcar nas fotos o local onde ela foi tirada, assim como citar algum endereço em seu perfil.

A possibilidade do compartilhamento da localização estreita a distância entre o mundo físico e o mundo online dos serviços de rede social, com isto são criadas novas oportunidades e desafios relacionados a sistemas de recomendação.

Uma localização é um novo objeto em uma rede social baseada em localização, gerando novas relações entre usuários, entre localizações e entre usuários e localizações [2]. Pode-se haver agora diversos casos de recomendação, além do caso óbvio de recomendação de locais, é possível recomendar novos amigos, fotos, vídeos, tudo baseado em localizações. Com o histórico de localizações do usuário, podem ser aprendidos os seus gostos e costumes, e com

o aumento dos dados vindos do usuário é possível formar uma opinião social sobre tal localização.

1.2. Objetivo

O objetivo deste trabalho é analisar o contexto das redes sociais baseadas em localização, as técnicas de recomendação de localização nestas circunstâncias, implementar um algoritmo a ser testado em uma base de dados real e analisar os resultados.

1.3. Metodologia

Depois de escolher o tipo de sistema de recomendação, o primeiro passo foi fazer uma revisão de literatura. Isso foi necessário para o aprendizado de conceitos básicos, dos diferentes objetivos dentro da área, desafios e análise das técnicas relacionadas. Após a revisão literária, foi escolhido o artigo *User Preferences-Based and Time-Sensitive Location Recommendation Using Check-In* [3] a servir como guia. Baseado neste artigo, foi-se implementado o algoritmo, testado sobre uma base de dados real e feita a análise dos resultados.

1.4. Estrutura do Trabalho

Este trabalho é apresentado da seguinte maneira: o capítulo 2 faz uma revisão de literatura mostrando os conceitos básicos ao trabalhar com redes sociais baseadas em localização assim como os sistemas de localização associados e as diferentes categorias dentro deste contexto. O capítulo 3 mostra em detalhes e em partes o método proposto e a explicação de cada uma destas partes. Os experimentos são mostrados no capítulo 4, onde é descrito o dataset utilizado e a avaliação dos resultados. O capítulo 5 mostra a aprendizagem neste trabalho, discute o alcance do objetivo e cita caminhos para trabalhos futuros.

2. Revisão de Literatura

Este capítulo é responsável por explicar os conceitos básicos de sistemas de recomendações para redes sociais baseadas em localização. Seção 2.1 explica a adição do contexto de localização em tais redes e as propriedades destas enquanto que na seção 2.2 serão explicados os tipos de categorização nestas redes sociais.

2.1. Sobre redes sociais baseadas em localização

Redes sociais são representações digitais da vida real do ciclo social das pessoas, nestas os usuários se conectam com amigos, compartilhando momentos de sua vida, pensamentos, conhecimento e outros por meio dos mais variados tipos de compartilhamento (música, vídeo, texto, etc). Além de se conectar a amigos, as redes sociais possuem grupos onde o usuário pode encontrar pessoas com interesses em comum.

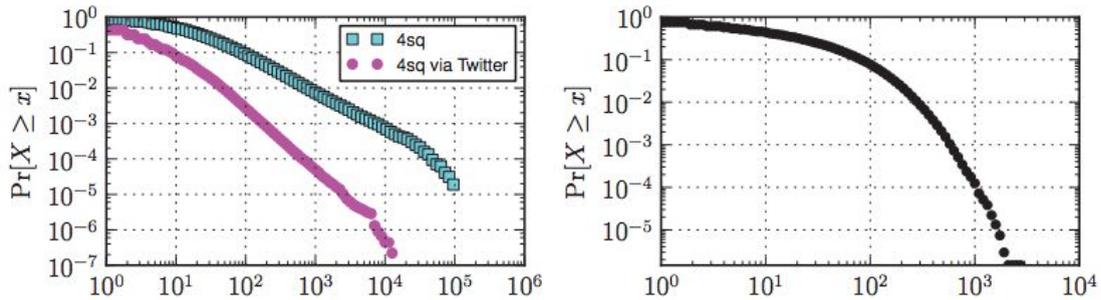
Redes sociais baseadas em localização ou Redes geográficas sociais é formalmente definida como o tipo de rede social aonde serviços geográficos e funções como geocoding e geotagging são utilizadas para capacitar dinâmicas sociais adicionais [4]. A adição de contexto geográfico as redes sociais torna a representação social digital mais próxima do conjunto social na vida real.

2.2. Contexto e compreensão da adição de localização às redes sociais

A adição do contexto de geolocalização as redes sociais dá a possibilidade de coleta de muitos dados a respeito dos interesses do usuário, quanto mais check-ins ou compartilhamento de localização mais evidências se tem para traçar os seus interesses.

Além da caracterização do usuário, é possível também encontrar padrões e compreender os comportamentos do mesmo. Em [5] foram coletados aproximadamente 12,000,000 check-ins de mais de 679,000 usuários em um período de 111 dias, e foi possível ver alguns padrões de comportamento:

Os usuários tendem a seguir a lei da potência. Visitam mais locais que são próximos com mais frequência do que locais distantes, na Figura 1 mostra-se o número de check-ins por localização adquirido no Twitter e o número de check-ins reportados no Foursquare, em ambos é visível a tendência em respeitar a lei da potência [5]:

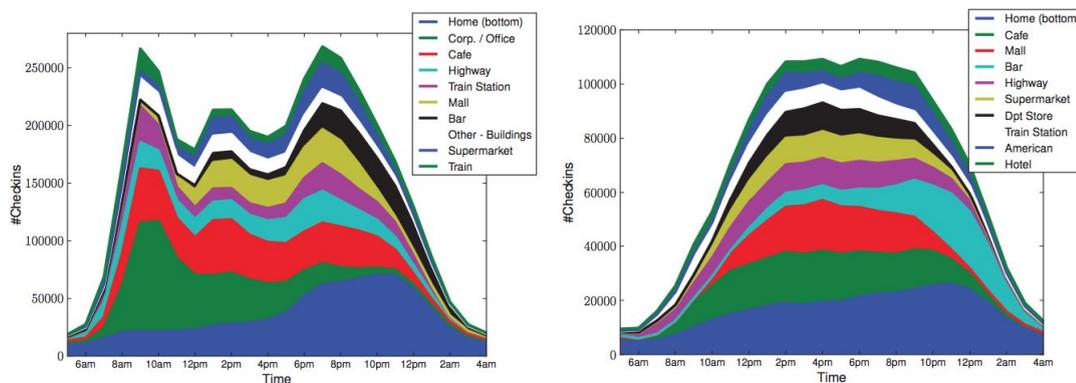


(a) Check-ins do lugar CCDF

(b) Check-ins do usuário CCDF

Figura 1. Funções Distributivas Acumulativas Complementares para o número de check-ins nos lugares (esquerda) e o número de check-ins por usuários (direita) [5].

Outra tendência apresentada em [5] é que em dias úteis o padrão dos usuários apresentam picos no início do trabalho, durante a hora do almoço e na saída do trabalho, enquanto no fim de semana apresenta-se um platô entre 12pm e 10pm (ver Figura 2).



(a) Dias úteis

(b) Final de semana

Figura 2. Plot empilhado das 10 categorias mais populares sobre os dias úteis e os fins de semana. A popularidade decresce de cima para baixo [5].

2.3. Propriedades das redes sociais baseadas em localização

Em uma rede social baseada em geolocalização, existem três importantes propriedades (ver figura 3):

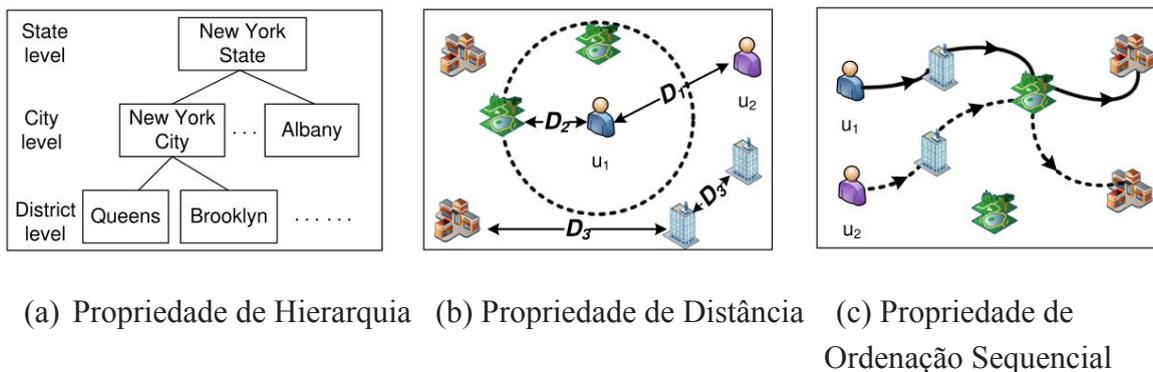


Figura 3. Propriedades nas redes sociais baseadas em localização [2].

- Hierarquia, que também pode ser chamada de escala, define a granularidade de uma localização, um exemplo seria um bar que está em um bairro e este está em uma cidade, neste exemplo temos 3 lugares, cada um com granularidade diferente. A importância disto é dado que um usuário frequenta um local com granularidade baixa (o bar), a chance dele conhecer alguém no local que o bar se encontra com granularidade mais alta (o bairro), é maior.
- Distâncias, estas podem ser medidas entre diferentes usuários, entre um usuário e um local e entre locais. A distância afeta a rede social de três maneiras: 1) A distância entre usuários influencia em sua similaridade, usuários que moram muito próximos são mais propensos a serem amigos [6]. 2) A distância entre local e usuário irá influenciar no nível de interesse que o usuário tem de visitá-lo, usuários no Foursquare visitam restaurantes mais próximos a suas casas com mais frequência do que outros [7]. A distância entre localizações influencia a similaridade entre elas, por exemplo, é comum existir restaurantes próximos a outros.

- Ordenação Sequencial, Pode-se criar uma ordem cronológica de ordem quando visto um padrão de visitas do usuário, um exemplo é o usuário que vai pra faculdade e logo depois vai para a academia de segunda a sexta, a probabilidade deste usuário ir pra academia depois da faculdade será então muito alta.

2.4. Categorizações dos sistemas de recomendação

Nesta subseção serão apresentadas as categorias relacionadas aos sistemas de recomendação para redes sociais baseadas em localização. Iremos categorizar por objetivo, metodologia e fonte de dados (ver figura 4).

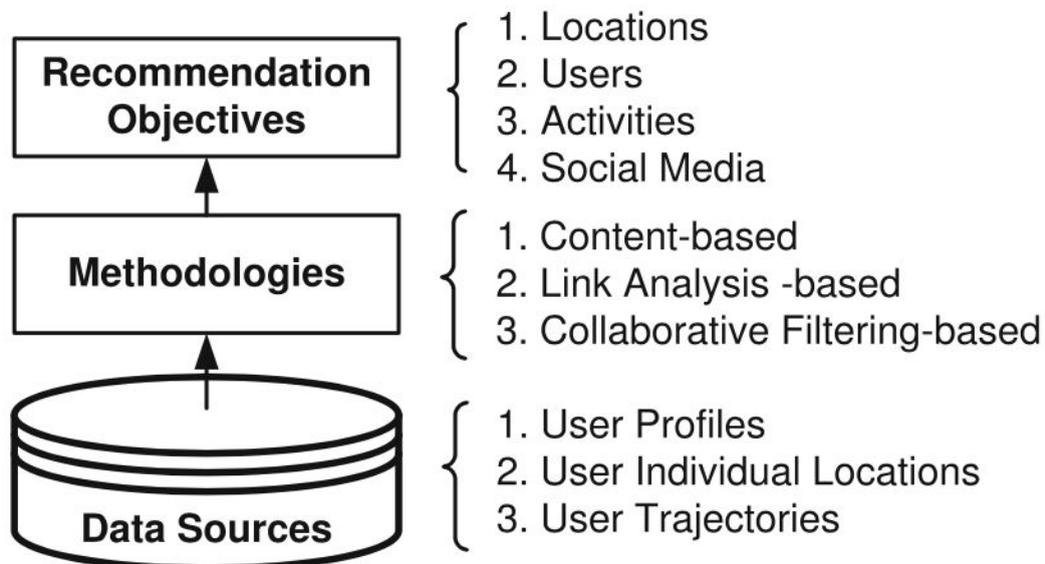


Figura 4. Visão geral sobre as categorias de sistemas de recomendação para redes sociais baseadas em localização [2].

2.4.1. Categorização dos sistemas de recomendação por objetivo

Os sistemas de recomendação de localização nas redes sociais baseadas em localização podem ser categorizados de quatro formas: 1) Recomendação de localização única (como exemplo bar) para o usuário. 2) Recomendação de usuário, baseado na similaridade entre um usuário e outro (como frequentar os mesmos locais ou morarem muito próximos) recomenda-o como amigo, também pode ser útil para recomendar comunidades com os

Algoritmos de análise de links são muito utilizados para ranquear páginas web. Em uma rede social baseada em geolocalização, esses algoritmos servem para encontrar nós que possuem grande influência sobre outros. Em [10] o autor estende o algoritmo Hypertext Induced Topic Search (HITS) para extrair locais de grande importância (muito visitados) e usuários experientes (que viajam muito), neste algoritmo tem-se hubs que servem como um catálogo para os nós authorities, bons hubs apontam para vários nós enquanto bons authorities são apontados por vários nós (ver figura 6). As vantagens deste algoritmo é que as experiências dos usuários são levadas em consideração e não há o problema do cold start [8]. Uma desvantagem é que as recomendações são superficiais em relação aos gostos pessoais de cada usuário.

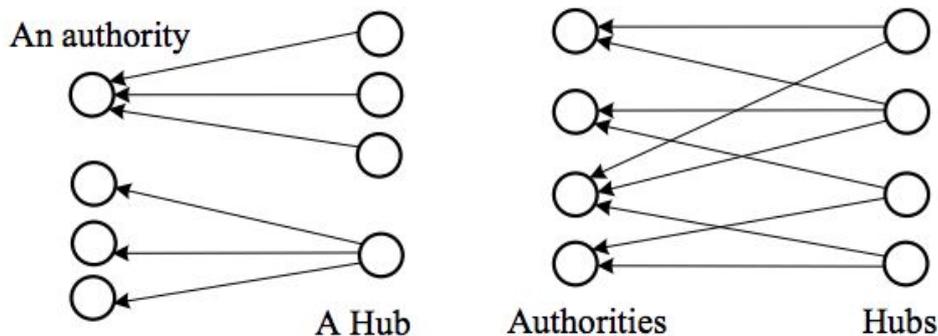


Figura 6. O conceito básico do modelo HITS [10].

Recomendação baseada em filtragem colaborativa

Filtragem colaborativa parte-se do princípio que um usuário irá gostar de visitar lugares muito visitados por usuários similares a ele. As vantagens desta abordagem são: não é necessária uma informação bem estruturada dos usuários ou objetos e é levada em consideração a opinião da comunidade (o que provê muita qualidade). As desvantagens são: quando há pouca informação o modelo falha em fazer recomendações de qualidade, dependendo do número de itens pode-se demorar muito para concluir o algoritmo na execução do programa e não lida bem com o problema de cold start [8]. Esta abordagem consiste de três etapas:

- Seleção de candidatos: nesta etapa escolhe-se um subgrupo de nós candidatos, nos algoritmos comuns se utiliza os usuários mais parecidos (podem também ser locais ou

atividades) como candidatos, além disso também pode ser utilizado algum tipo de limite geográfico para a seleção dos candidatos, como feito em [11] onde se levava em consideração a distância dos possíveis candidatos até onde foi feita a query.

- Inferência de similaridade: nas redes sociais baseadas em localização a similaridade entre dois usuários é encontrada pelo histórico de localizações (localizações em comum) e ranqueamento do usuário. Pode-se categorizar os modelos em filtragem colaborativa em dois grupos: o que mede a similaridade entre cada par de usuário e que mede a similaridade entre cada par de itens (como música, vídeos, etc).
- Predição de pontuação de recomendação: É dada uma pontuação para cada objeto do grupo candidato, é utilizado o ranqueamento dos usuários e a similaridade entre eles.

2.4.3. Categorização dos sistemas de recomendação por fonte de dados

Podemos categorizar os sistemas de recomendação por fonte de dados de três maneiras:

- Perfis do usuário: muito utilizado em sistemas de recomendação baseado em conteúdo, como em [9], esta fonte de dados fornece dados estruturados que informam os interesses, informação geográfica, compartilhamentos e outros dados que caracterizam o usuário.
- Histórico online do usuário: pode ser coletado por ranqueamento feito pelo usuário, histórico de pesquisa e padrões de interação do usuário. No caso de ranqueamento, pode ser utilizado para encontrar usuários parecidos e locais. Nos padrões de interação veremos o que o usuário pode ter gostado, tags que utiliza, comentários e etc. Os padrões do usuário são utilizados para a descoberta de grupos que o usuário tem em comum, sugestão de amizade para outros usuários, etc.
- Histórico de locais do usuário: o histórico dos check-ins do usuário são o seu rastro digital e dá a possibilidade de ter uma informação mais aprimorada sobre seu comportamento. Além disso é possível recomendar usuários como amigos a partir das interseções (das trajetórias) entre dois usuários.

3. Método Proposto

Nesta seção será discutido o método proposto que serviu como guia para os experimentos, irá primeiramente ser mostrado os fatores incluídos na construção do algoritmo e em seguida a construção do algoritmo. O código algoritmo se encontra em [15].

3.1. Recomendação sensível ao tempo e baseada em preferência de usuário.

O método proposto [3] consiste na construção de um sistema que leva em consideração quatro fatores: 1) a popularidade da localização; 2) a preferência do usuário; 3) a influência social dos amigos do usuário e amigos que deram check-in na mesma localização e 4) a hora da visita do usuário.

A hora da visita do usuário é de grande importância pois os locais tendem a variar bastante em número de visitantes de acordo com a hora do dia. A Figura 7 representa a distribuição de check-ins em dois locais diferentes, de acordo com com essa representação é notável a diferença da variação de visitas nos dois locais. O sistema de recomendação deve atentar a isto e levar em consideração para pontuar de maneira eficaz. Por exemplo, dado que são 21:00, o local da figura 7b deveria ser mais propenso a ser recomendado do que o da figura 7a.

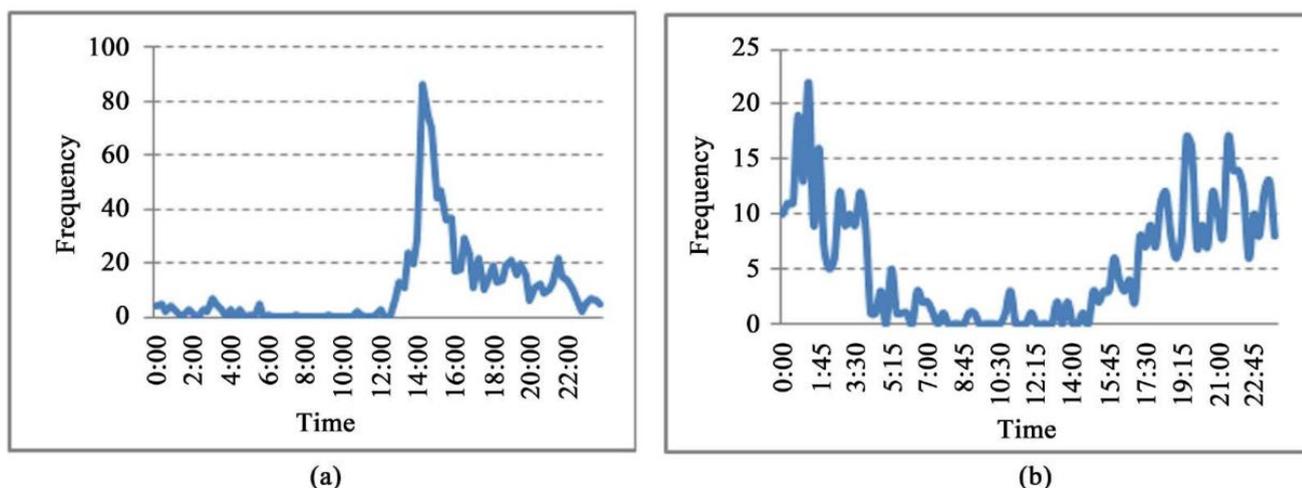


Figura 7. Distribuição de check-ins da (a) localização 153.505, (b) 14.710 [3].

Além da distribuição de check-ins das localizações, os padrões dos dos usuários também são de grande importância na construção de um sistema de recomendação, na Figura 8 temos um exemplo de dois usuários, em algumas localizações estes fazem pouquíssimos check-ins, entretanto em outras a frequência é muito maior.

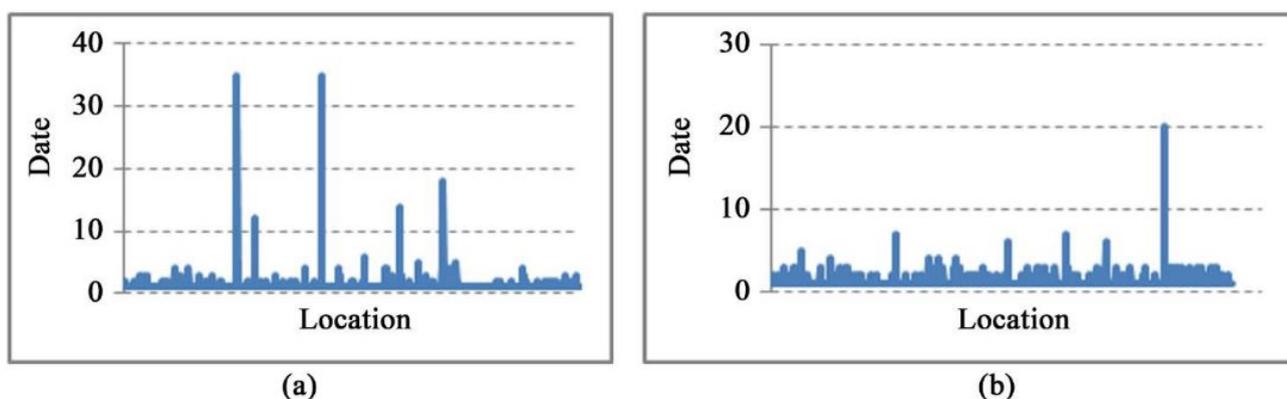


Figura 8. O número de datas que o (a) usuário 1251 visitou e (b) o usuário 2 visitou [3].

3.2. Fatores considerados no sistema de recomendação.

- Popularidade da localização: uma localização será muito popular se costuma receber muitas visitas de usuários, logo, esta tem a capacidade de despertar o interesse de outros usuários a visitá-la, esta popularidade pode ser medida facilmente através do número de usuários que deram check-in nesta localização.
- Preferência do usuário: apenas a popularidade da localização não bastaria, pois a recomendação seria muito genérica. Imaginemos que um usuário frequente determinada localização em determinado horário e que esta não seja muito popular, as chances deste usuário visitar a mesma localização é bem maior que ele visitar uma mais popular no mesmo horário. Para analisar e inferir a preferência do usuário é utilizado o seu histórico de check-ins.
- O momento da visita do usuário: a regularidade de visitas dos usuários é bastante pessoal e varia bastante de um para outro, por tanto, isto deve ser levado em

consideração. Para isto é usado o tempo corrente com a característica temporal do local.

- Os interesses dos amigos: é muito provável que se amigos de um usuário visitam com muita frequência uma localização, este usuário irá provavelmente se interessar em visitar o local. Para inferir este interesse, é preciso checar o histórico de check-ins dos amigos do usuário.

3.3. Modelos de recomendação de localização

3.3.1. Características temporais e popularidade do local

Quanto mais popular o local, mais significa que este possui alta qualidade em seus serviços. A característica de popularidade significa que vários usuários visitam aquela localização com muita frequência, logo, as chances daquele local ser do interesse de outros usuários é muito alta. em [3] a popularidade de uma localização l_i a ser recomendada ao usuário u_k é definida da seguinte maneira, onde U e L são os grupos de usuários e localizações e $c_{u,j}$ é a quantidade de check-ins na localização l_j :

$$pop_m(u_k, l_i) = \frac{\sum_{u \in U} c_{u,i}}{\sum_{u \in U} \sum_{l_j \in L} c_{u,j}}$$

Se uma localização é muito frequentada por um ou poucos usuários e estes sempre fazem check-in, teremos a falsa impressão de que este local é popular, para evitar este problema, [3] sugere utilizar entropia para o cálculo da popularidade do local:

$$pop_e(u_k, l_i) = - \sum_{u \in U} \frac{c_{u,i}}{\sum_{u \in U} c_{u,i}} \log \frac{c_{u,i}}{\sum_{u \in U} c_{u,i}}$$

Como discutido na seção 3.1, devemos estar atentos ao horário corrente para saber qual localização recomendar, pois estas podem ser muito populares em alguns momentos e pouco populares em outros. A equação então é alterada para a seguinte, sendo $pop(u_k, l_i) \in \{pop_m(u_k, l_i), pop_e(u_k, l_i)\}$, t_u é o momento em que o check-in ocorreu e t o momento corrente do usuário; de acordo com a natureza exponencial $e^{-|t_u-t|}$ sabemos que quanto mais próximo ao horário de visita de usuário, maior a similaridade entre eles [3]:

$$pop_t(u_k, l_i) = \frac{\sum_{u \in U} e^{-|t_u-t|} c_{u,i}}{\sum_{u \in U} c_{u,i}} \cdot p(u_k, l_i)$$

3.3.2. Preferência do usuário

Como discutido na seção 3.1, a diferença na quantidade e frequência de check-ins entre dois usuários pode ser bastante diferente, por isso devemos levar em consideração a preferência individual de cada usuário. Implica-se está preferência através do histórico de check-ins, em [3] é usado TF-IDF para essa implicação, onde U_i é o grupo de usuários que deram check-in na localização l_i :

$$pre(u_k, l_i) = \frac{c_{u_k,i}}{\sum_{l_j \in L} c_{u_k,j}} \cdot \log \frac{\sum_{u \in U} c_{u,i}}{|U_i|}$$

Caso um usuário foi ao mesmo lugar várias vezes em um espaço de tempo curto, não podemos determinar se ele realmente se interessa por esse local, entretanto, se ele foi ao local em vários dias, podemos dizer que há interesse, [3] altera a equação anterior adicionando um fator relacionado aos dias que o usuário foi ao local, onde $d_{u_k,i}$ é o número de dias que o usuário u_k fez check-in na localização l_i :

$$pre_i(u_k, l_i) = \frac{c_{u_k,i}}{\sum_{l_j \in L} c_{u_k,j}} \cdot \frac{d_{u_k,i}}{\sum_{l_j \in L} d_{u_k,j}} \cdot \log \frac{\sum_{u \in U} c_{u,i}}{|U_i|}$$

3.3.3. Influência Social dos amigos

Amigos de um usuário tem grandes chances de frequentarem os mesmos lugares, talvez até com a mesma frequência, em [3] é inferida a similaridade entre dois usuários da seguinte maneira, sendo F_u o grupo de amigos do usuário u , desta maneira é encontrado o número de amigos em comum entre os dois usuários e é escalado pelo número de amigos do usuário que possui menos amigos:

$$M_f(u, v) = \frac{|F_u \cap F_v|}{\min(|F_u|, |F_v|)}$$

Alternativamente [3] calcula a similaridade entre dois amigos pelos lugares que já visitou, dessa maneira é mostrado se os usuários realmente tem as mesmas preferências (L_u é o grupo de check-ins do usuário u):

$$M_l(u, v) = \frac{|L_u \cap L_v|}{\min(|L_u|, |L_v|)}$$

Em [12] é medida a similaridade entre dois usuários pelos itens compartilhados entre eles [3]. Em [3] é utilizado esta medida para inferir a similaridade entre dois usuários que deram check-in no mesmo local calculando-se pelo inverso da frequência logarítmica de seus check-ins em comum:

$$S(u, v) = \sum_{l_j \in L_u \cap L_v} \frac{1}{\log |U_j|}$$

Recomendação baseada em amigos utiliza algoritmo de filtragem colaborativa [13], onde $w_{v,u} \in \{M_f(u, v), M_l(u, v), S(u, v)\}$:

$$sim(u_k, l_i) = \frac{\sum_{v \in F_u} w_{v,u} \cdot c_{v,i}}{\sum_{v \in S_u} w_{v,u}}$$

3.3.4. Modelo de Fusão

Em [3] chega-se no modelo de fusão onde serão utilizados os modelos mencionados nas seções 3.3.1, 3.3.2 e 3.3.3, onde α e β estão no intervalo $[0, 1]$ e $\alpha + \beta = 1$:

$$RS(u_k, l_i) = (1 - \alpha - \beta) pop(u_k, l_i) + \alpha sim(u_k, l_i) + \beta pre(u_k, l_i)$$

4. Experimentos

4.1. Dados

Foram utilizados os dados da antiga rede social baseada em localização Brightkite [14], os dados contém 4.491.143 check-ins, 58.228 usuários e 214.078 arestas entre os usuários. Para o experimento, devida à complexidade no tempo do algoritmo, foram utilizados apenas os 500 primeiros usuários (estes ordenados pelo seu id) resultando em 255.414 check-ins. Destes 500 usuários, foram selecionados aleatoriamente 100 usuários e pego o último check-in de cada um deles para realizar os testes.

4.2. Avaliação

Para avaliarmos os resultados, aplicamos o algoritmo em cada local, para cada usuário e então teremos os locais ranqueados pelo seu score. A partir disto calcularemos a média de precision e recall calculados em todos os testes. Precision é calculado sendo $(\text{número de acertos})/N$, que neste caso *número de acertos* será sempre no máximo um, enquanto N é o número de locais no topo do ranking, teremos um *acerto* quando o local for igual ao local em que o usuário fez check-in. Para o cálculo do Recall teremos 100% se o resultado aparece até a posição N. Finalmente é calculado o f-measure, que é a medida da precisão de um teste [16], calculado pela fórmula $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$.

Primeiro foi necessário avaliar qual seria o melhor algoritmo para fazer a recomendação, utilizando $\alpha=0.333$ e $\beta=0.333$, foram testados os pares de possibilidades dados pelos conjuntos, no caso de popularidade do local tem-se $pop(u_k, l_i) \in \{pop_m(u_k, l_i), pop_e(u_k, l_i)\}$ e no caso da similaridade tem-se $w_{v,u} \in \{M_f(u, v), M_l(u, v), S(u, v)\}$, logo, temos 6 pares a serem testados, serão calculados precision e recall para estimar a qualidade do algoritmo. Os testes foram com N=1 e

tiveram-se os resultados apresentados na tabela 1 em termos de precision/recall e na tabela 2 em termos de f-measure:

Tabela 1. Resultado em precision/recall para cada par de possibilidades, com $N=1$.

| $pop(u_k, l_i) / w_{v,u}$ | $M_f(u, v)$ | $M_l(u, v)$ | $S(u, v)$ |
|---------------------------|-------------|-------------|-------------|
| $pop_m(u_k, l_i)$ | 0.12 / 0.12 | 0.08 / 0.08 | 0.08 / 0.08 |
| $pop_e(u_k, l_i)$ | 0.04 / 0.04 | 0.08 / 0.08 | 0.04 / 0.04 |

Tabela 2. Resultado em f-measure para cada par de possibilidades, com $N=1$.

| $pop(u_k, l_i) / w_{v,u}$ | $M_f(u, v)$ | $M_l(u, v)$ | $S(u, v)$ |
|---------------------------|-------------|-------------|-----------|
| $pop_m(u_k, l_i)$ | 0.12 | 0.08 | 0.08 |
| $pop_e(u_k, l_i)$ | 0.04 | 0.08 | 0.04 |

Escolhido o melhor par para o algoritmo, que no caso foi o $(pop_m(u_k, l_i)$ e $M_f(u, v)$), foram feitas variações no α e β , sendo estas dentro do conjunto $\{0, 0.33, 0.5, 1\}$, onde 0.33 inclui todos os modelos de maneira equivalente; 0.5 e 0 utiliza apenas dois modelos e 1 e 0 utiliza apenas um modelo. Os resultados podem ser vistos nas Tabelas 3 e 4 em termos de precision/recall e f-measure (as células com "-" indicam um par (α, β) que não foi testado):

Tabela 3. Resultado em precision/recall para cada par de possibilidades α e $\beta \in \{0, 0.33, 0.5, 1\}$.

| α / β | 0 | 0.333 | 0.5 | 1 |
|------------------|-------------|-------------|-------------|-----------|
| 0 | 0.0 / 0.0 | - | 0.04 / 0.04 | 0.2 / 0.2 |
| 0.333 | - | 0.12 / 0.12 | - | - |
| 0.5 | 0.12 / 0.12 | - | 0.16 / 0.16 | - |
| 1 | 0.12 / 0.12 | - | - | - |

Tabela 4. Resultado em f-measure para cada par de possibilidades α e $\beta \in \{0, 0.33, 0.5, 1\}$.

| α / β | 0 | 0.333 | 0.5 | 1 |
|------------------|------|-------|------|-----|
| 0 | 0 | - | 0.04 | 0.2 |
| 0.333 | - | 0.12 | - | - |
| 0.5 | 0.12 | - | 0.16 | - |
| 1 | 0.12 | - | - | - |

Como é visível na Tabela 2, a melhor combinação foi com $\alpha = 0$ e $\beta = 1$, ou seja, quando apenas é levando em conta a preferência de usuário o algoritmo demonstrou ser mais efetivo e quando levou-se em conta apenas a popularidade do local o algoritmo se mostrou menos efetivo.

4.3. Resultados

No experimento deste artigo obtiveram-se utilizando os coeficientes $\alpha = 0$ e $\beta = 1$, os resultados mostrados na Figura 9 para precision, Figura 10 para recall e Figura 11 para f-measure.

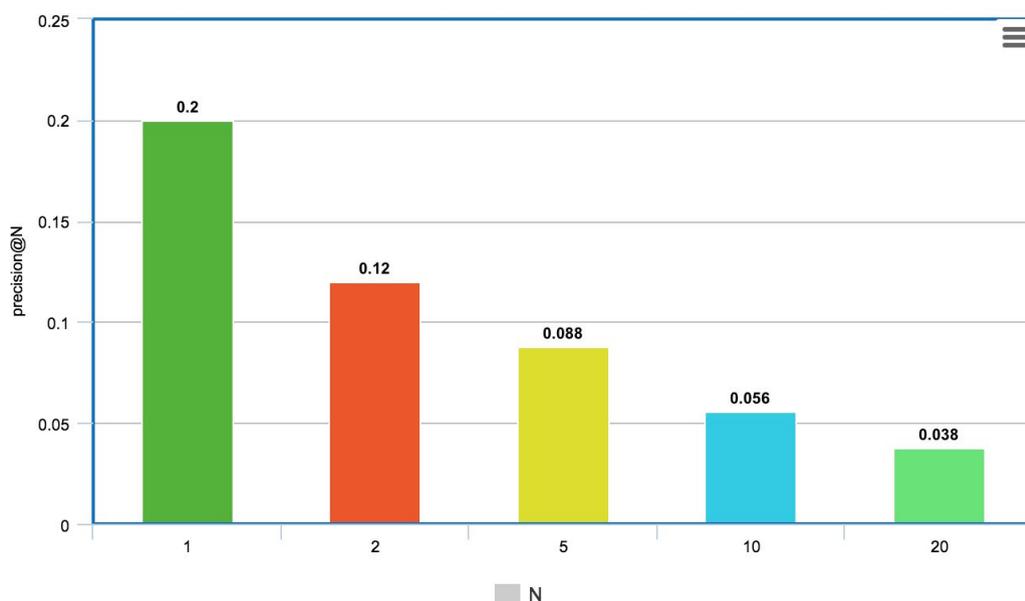


Figura 9. Precision utilizando $\alpha = 0$ e $\beta = 1$.

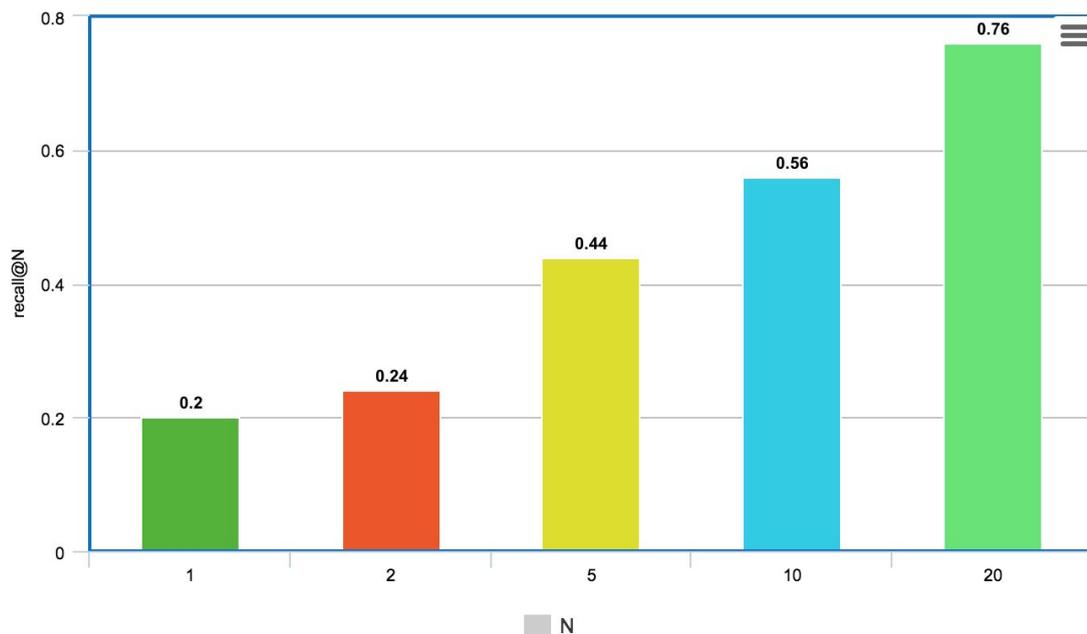


Figura 10. Recall utilizando $\alpha = 0$ e $\beta = 1$.

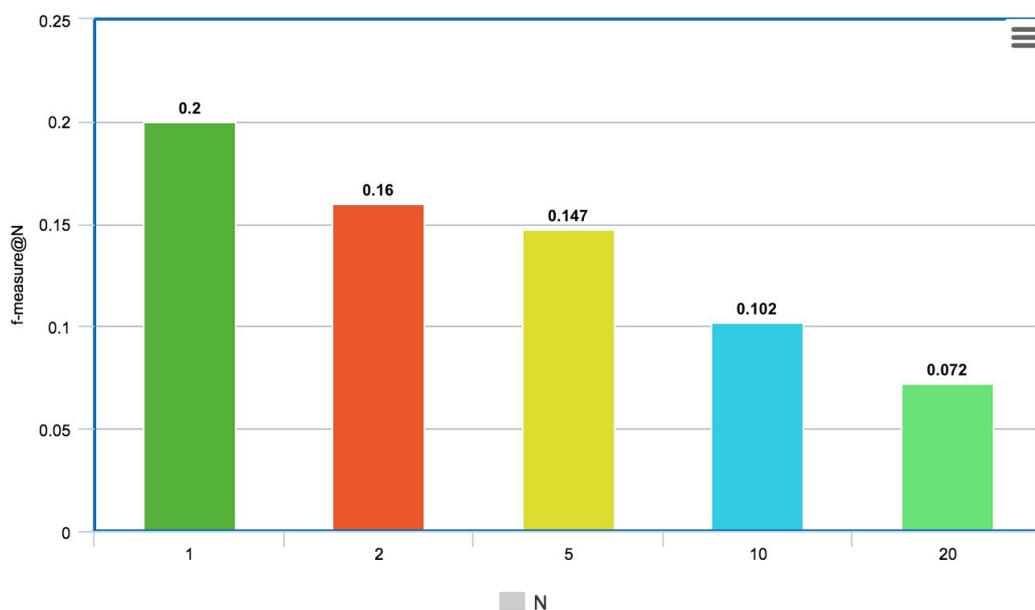


Figura 11. F-measure utilizando $\alpha = 0$ e $\beta = 1$.

Comparando com o experimento feito em [3], tivemos uma queda no Precision e um aumento no Recall, entretanto o algoritmo ainda se mostrou superior sobre os outros, de acordo com o experimento em [3], os parâmetros que deram os melhores resultados foram $\alpha = 0$ e $\beta = 1$. Mostrou-se então que utilizando apenas o modelo de preferência de usuário sozinho, obteve-se uma superioridade em relação aos outros algoritmos pois o histórico de check-ins do usuário trás um grande peso sobre seus interesses e hábitos.

5. Conclusão

O intuito deste trabalho foi fazer um estudo sobre sistemas de recomendação de localização para redes sociais baseadas em localização e a partir deste estudo selecionar um algoritmo para fazer uma replicação do experimento. O artigo selecionado foi o [3] e o experimento foi realizado com outro dataset; a amostra foi menor, entretanto foram comparados os resultados na amostra variando os coeficientes do modelo de fusão. O resultado mostrou que o modelo quando levado em consideração apenas a popularidade do local é menos eficiente e o modelo de preferência de usuário sozinho conseguiu ser melhor que as outras variações de α e β e que este comparado ao experimento feito em [3] mostrou uma menor Precision e um maior Recall, mas ainda assim, superior aos outros algoritmos.

Em trabalhos futuros poderia ser utilizado uma quantidade maior de dados de teste e fazer uma variação maior dos parâmetros α e β a fim de encontrar pesos para cada modelo que tragam os melhores resultados.

Referências

- [1] Panagiotis S., Alexis P., Yannis M., Pinar S. e Ismail T. Geo-social recommendations based on incremental tensor reduction and local path traversal. Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, 2011.
- [2] J. Bao, Y. Zheng, D. Wilkie, e M. F. Mokbel. A survey on recommendations in location-based social networks. ACM TIST. V, N, Article A, 2014.
- [3] Zhang, S. e Ren, K. User Preferences-Based and Time-Sensitive Location Recommendation Using Check-In Data. Journal of Computer and Communications, 2015.
- [4] Quercia D., Lathia N., Calabrese F., Di Lorenzo G. e Crowcroft J. Recommending social events from mobile phone location data. In: International conference on data mining. IEEE International Conference on Data Mining, 2010.
- [5] Noulas A., Scellato S., Mascolo C., Pontil M. An empirical study of geographic user activity patterns in foursquare. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.
- [6] DeScioli P., Kurzban R., Koch EN., Liben-Nowell D. Best friends alliances, friend ranking, and the myspace social network. Perspectives on Psychological Science, 2011.
- [7] 55. Levandoski J., Sarwat M., Eldawy A., Mokbel M. Lars: a location-aware recommender system. IEEE 28th International Conference on Data Engineering, ICDE 2012, 2012.
- [8] The Cold Start Problem for Recommender Systems,
<http://www.yusp.com/blog/cold-start-problem-recommender-systems/>
- [9] Park M-H, Hong J-H, Cho S-B. Location-based recommendation system using bayesian users preference model in mobile devices. In: Ubiquitous intelligence and computing. International Conference on Ubiquitous Intelligence and Computing, 2007.
- [10] Zheng Y., Zhang L., Xie X., MaW-Y. Mining interesting locations and travel sequences from gps trajectories. Proceedings of the 18th international conference on World wide web, 2009.

- [11] Chow C-Y, Bao J., Mokbel MF. Towards location-based social networking services. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, LBSN, 2010.
- [12] Adamic, L.A. and Adar, E. Friends and Neighbors on the Web. Social Networks. Journal of Computer and Communications, 2003.
- [13] Ye, M., Yin, P., Lee, W.C. and Lee, D.L. Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation. SIGIR, 2011.
- [14] E. Cho, S. A. Myers, J. Leskovec. Friendship and Mobility: Friendship and Mobility: User Movement in Location-Based Social Networks. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011.
- [15] Código do algoritmo, <https://github.com/vapp94/loc-recommendation>.
- [16] F1 score, https://en.wikipedia.org/wiki/F1_score.