

Explorando o nível de atividade do usuário para melhorar a precisão dos sistemas de recomendação de pontos de interesse*

Luiz Chaves¹, Nícollas Silva², Leonardo Rocha¹

¹ DCOMP/UFSJ - São João del-Rei, MG, Brasil

² DCC/UFMG - Belo Horizonte, MG, Brasil

luizfufsj@gmail.com, lcrocha@ufsj.edu.br, ncsilvaa@dcc.ufmg.br

Abstract. *Recommender Systems (RS) have been applied in several scenarios, such as entertainment, e-commerce, and nowadays, in Location-Based Social Network (LBSN) to recommend points-of-interest (POIs). In POIs domains, there is still an opportunity for improvements, on which it is necessary to consider the geographic influence of them. The actual main proposals are not able to achieve satisfactory results. In this work, we open a new research perspective, proposing a post-processing approach that can be used with any RS. We measure the activity level of users in different subareas of a city and use it to re-order the POIs retrieved by an RS. We evaluate our proposal considering six recommender systems and three datasets from Yelp achieving gains up to 15% of precision.*

Resumo. *Sistemas de Recomendação (SsR) são aplicados em diversos cenários, tais como comércio eletrônico e, atualmente, em Redes Sociais Baseadas em Localização para recomendar pontos de interesse (POIs). Para cenário de POIs é necessário considerar a influência geográfica deles. As propostas atuais não alcançam resultados satisfatórios. Neste trabalho, abrimos uma nova perspectiva de pesquisa, propondo uma abordagem de pós-processamento que pode ser usada com qualquer SR. Medimos o nível de atividade dos usuários em diferentes subáreas de uma cidade e o usamos para reordenar os POIs recuperados. Avaliamos nossa proposta considerando seis SsR e três conjuntos de dados do Yelp, obtendo ganhos de até 15% de precisão.*

1. Introdução

Aplicações Web têm investido em Sistemas de Recomendação (SsR) com o objetivo de satisfazer seus usuários. Em aplicações como a Amazon.com, a tarefa do SR é apresentar produtos de interesse dos usuários [Bobadilla et al. 2013]. No caso das redes sociais, como o Facebook, a tarefa é identificar as publicações mais relevantes para os usuários [Kumar et al. 2015]. Por sua vez, nas redes sociais baseadas em localização (i.e., *Location Based Social Network* - LBSN), tais como *Foursquare*, *Gowalla* e *Yelp*, a tarefa dos SsR é identificar os interesses dos usuários em pontos de interesse (i.e., *Points of Interest* - POIs) como restaurantes, bares, cinemas, teatros, museus, e outros [Lu et al. 2015].

Em domínios tradicionais, tais como comércio eletrônico (produtos) e entretenimento (filmes e músicas), existem diversas propostas de SsR capazes de atingir seus objetivos [Bobadilla et al. 2013, Silva et al. 2019]. Contudo, a recomendação de POIs ainda é um desafio em aberto devido às suas características peculiares. Diferente dos outros

*Esse trabalho foi parcialmente financiado por CNPq, CAPES, FINER, Fapemig, e INWEB.

cenários, onde *consumir* um item significa que o usuário apenas assistiu ou comprou um determinado item, em POIs consumir significa que o usuário teve que se deslocar até uma determinada região para visitá-lo. Ou seja, existe um fator geográfico intrinsecamente relacionado as atividades dos usuários que deve ser levado em conta [Ye et al. 2011, Liu et al. 2014]. Assim, os principais trabalhos em recomendação de POIs se preocupam em obedecer a Primeira Lei da Geografia de Tobler de que “*tudo está relacionado entre si, mas coisas próximas são mais relacionadas que coisas distantes*” [Miller 2004].

As propostas encontradas na literatura consistem de adaptações de SsR tradicionais para o cenário de recomendação de POIs [Yu and Chen 2015], bem como alguns novos SsR especificamente propostos para esse problema [Ye et al. 2011, Lian et al. 2014, Zhang and Chow 2015]. Trata-se de um cenário de pesquisa relativamente novo e nenhuma dessas abordagens foram capazes de alcançar resultados efetivamente satisfatórios [Yu and Chen 2015]. Ao invés de propor um novo modelo de recomendação, propomos uma nova abordagem para auxiliar os SsR já existentes a identificarem POIs potencialmente relevantes aos usuários. Trata-se de uma etapa de pós-processamento que pode ser utilizada em conjunto com qualquer SR, desde os tradicionais até aqueles específicos para POI. Nossa proposta consiste em medir o nível de atividade de cada usuário nas diversas sub-regiões que compõem uma determinada área (e.g., cidade) e, a partir desse nível de atividade, ponderar a lista de recomendação gerada por algum SR de acordo com esses níveis de atividade. Assim, garantimos que POIs pertencentes à sub-regiões de muita atividade do usuário recebem uma ponderação maior do que POIs pertencentes a sub-regiões de menor atividade. A partir dessa ponderação, essa lista é então reordenada e apresentada ao usuário.

Para avaliar nossa abordagem, aplicamos a etapa de pós-processamento proposta, considerando SsR usados em aplicações tradicionais (i.e., MostPopular, User-KNN e WRMF) e modelos específicos para o problema de recomendação de POIs (i.e., USG, GeoMF e GeoSoCa). Nessa avaliação, consideramos os dados da base *Yelp*, muito utilizada no cenário de POI, selecionando três cidades (i.e., Las Vegas, Phoenix e Charlotte). Os resultados mostram que, ainda que considerando um métrica simples para se medir o nível de atividade dos usuários, nossa estratégia é capaz de melhorar a qualidade da recomendação na grande maioria dos casos analisados, sendo alguns deles acima de 15% de ganho de precisão, como é o caso do algoritmo MostPopular para a cidade de Las Vegas. Até mesmo quando nossa etapa de pós-processamento é aplicada em SsR específicos de POI, é possível observar ganhos de qualidade, sendo alguns deles acima de 8% na precisão, como o GeoMF na cidade de Charlotte.

A principal contribuição desse trabalho é uma nova perspectiva de se abordar a recomendação de POIs, por meio de uma etapa de pós-processamento que pode ser utilizada por SsR. Essa estratégia abre uma nova linha de pesquisa, em que melhorias de nossa proposta podem ser avaliadas (i.e., estratégias para medir o nível de atividade dos usuários e/ou definir sub-regiões) ou novas etapas de pós-processamento propostas.

Todas as implementações necessárias para execução desse trabalho, bem como as execuções dos experimentos foram realizadas pelo aluno Luiz Felipe Chaves, sob a orientação do professor Leonardo Rocha. Toda a concepção do projeto bem como as análises de resultados foram feitas em conjunto, aluno e professor, com a colaboração do aluno de doutorado do Programa de Pós-Graduação do DCC/UFMG, Nícollas Silva.

2. Trabalhos Relacioandos

Dentro do domínio de recomendação, o cenário de POIs se difere dos demais devido às suas características peculiares. Primeiro, muitos trabalhos destacam que o problema de esparsidade é ainda maior no domínio de POIs, uma vez que o usuário enfrenta obstáculos físicos para visitar lugares [Ye et al. 2011, Liu et al. 2014]. Além disso, outros trabalhos destacam que a fonte de informação sobre os usuários difere dos demais. Nesse cenário, as principais fontes de informação para medir o interesse do usuário são (1) social; (2) temporal; e (3) geográfica. A influência social assume que as opiniões dos amigos presentes na rede social do usuário são mais relevantes que a dos não-amigos [Ye et al. 2010, Fernández-Tobías et al. 2016]. Por sua vez, a influência temporal refere-se ao fato de que os usuários tendem a visitar lugares, por exemplo restaurantes, em um período esporádico de tempo [Yuan et al. 2013, Liu et al. 2013]. Sobretudo, a influência geográfica é a principal característica a ser considerada pois ela é intrínseca desse cenário de recomendação, uma vez que os usuários e POIs estão situados em um espaço físico [Yu and Chen 2015].

Nesse sentido, embora existam diversas abordagens eficazes de recomendação nos cenários tradicionais, elas não são igualmente efetivas para esse caso. Por essa razão, encontramos na literatura trabalhos que exploram as diferentes características de recomendação de POIs. A influência temporal é explorada em [Gao et al. 2013], assumindo que usuários com históricos de consumo semelhantes, em um mesmo período de tempo, compartilham dos mesmos interesses. Por sua vez, a influência social é explorada em [Ye et al. 2011] e [Zhang and Chow 2015]. No trabalho de *Ye et al.* [Ye et al. 2011], a influência é extraída par a par entre os amigos de cada usuário. Em [Zhang and Chow 2015], é criada uma função de distribuição cumulativa para estimar as relações sociais. Por fim, a influência geográfica é considerada pela literatura como a mais importante e que apresenta os melhores resultados, uma vez que o usuário sempre levará em conta a distância de um local para visitá-lo [Ye et al. 2011, Zhang and Chow 2015, Lian et al. 2014, Li et al. 2015, Liu et al. 2014]. Dentre esses, podemos destacar os trabalhos apresentados por *Ye et al.* [Ye et al. 2011] e *Lian et al.* [Lian et al. 2014] como aqueles que são hoje considerados o estado da arte. Esses autores utilizam das informações passadas para avaliar a probabilidade de um usuário visitar um POI, em [Ye et al. 2011], ou uma região, dada a propagação da influência dos POIs no espaço geográfico [Lian et al. 2014]. Para ambos os casos, a probabilidade é considerada ainda no processo de construção do modelo de recomendação (i.e. na matriz de fatoração).

3. Filtragem Geográfica

A influência geográfica é um dos principais fatores que afetam as recomendações de POIs [Liu et al. 2014, Li et al. 2015]. Em geral, os usuários estão interessados em visitar lugares próximos de sua localização atual e/ou localizados nas regiões em que eles frequentam com maior frequência (alto nível de atividade) [Yu and Chen 2015]. Nossa proposta consiste em filtrar as recomendações apresentadas por qualquer modelo base, quantificando o nível de atividade do usuário em uma determinada região, para impulsionar ou penalizar determinados POIs recomendados. Primeiramente, dividimos uma região em subáreas com base nas sugestões da literatura [Han and Yamana 2017]. Em seguida, calculamos o nível de atividade do usuário em cada uma dessas sub-regiões. Por fim, utilizamos essas informações para processar as recomendações feitas a fim de melhorá-las. Denominamos essa etapa de pós-processamento de *Geo-Filtering*, a qual será detalhada nas seções seguintes.

3.1. Áreas e Subáreas de Recomendação

Inspirados pela abordagem de recomendação adotada por *Han & Yamana* [Han and Yamana 2017], inicialmente definimos uma área de recomendação A como uma cidade ou distrito delimitada por latitudes e longitudes, iniciais e finais. Dividimos essa mesma região A em subáreas quadradas de tamanho fixo e com dimensões de 0,5 x 0,5 km. Assim, criamos uma matriz de subáreas a_{ij} conforme podemos observar na Figura 1. Essa divisão fixa em 0,5km é um valor padrão que encontramos na literatura [Han and Yamana 2017], e onde os testes experimentais se mostraram mais eficazes.

0.6	0.8	0.7	0.4
0.5	0.9	0.4	0.1
0.2	0.5	0.9	0.2
0.1	0.8	0.7	0.3

Figura 1. Divisão em Subáreas

A ideia de divisão de A em subáreas é apresentada originalmente em *Han & Yamana* [Han and Yamana 2017] a fim de mitigar o problema da esparsidade dos POIs. Nesse caso, os autores propuseram um pré-processamento que filtra os POIs disponíveis para a recomendação, deixando apenas aqueles das subáreas mais movimentadas. Por sua vez, em nosso método, personalizamos essa definição, levando em conta a atividade do usuário em cada região. O principal objetivo dessa divisão é auxiliar nosso modelo a considerar a localização dos POIs e o nível de atividade dos usuários [Liu et al. 2013]. Apesar de simples, essa divisão em sub-regiões de tamanho fixo apresentou bons resultados. Entretanto, essa é uma das etapas de nossa estratégia que ainda pode ser melhorada. Por exemplo, estabelecer dimensões diferentes para cada sub-região de acordo com a quantidade de POIs presente ou até mesmo utilizar algoritmos de clusterização por densidade, como DBSCAN [A. et al. 2013], para estabelecer essas sub-regiões automaticamente.

3.2. Nível de Atividade do Usuário

Os principais trabalhos em recomendação de POIs se preocupam em obedecer a Primeira Lei da Geografia de Tobler. Baseados nessa lei, é possível derivar duas restrições fundamentais para atingir a satisfação dos usuários [Zhang and Chow 2015, Han and Yamana 2017]. Neste sentido, os usuários preferem: **(R1)** visitar lugares próximos de sua localização atual; **(R2)** visitar lugares localizados nas regiões em que eles possuem um alto nível de atividade.

Propomos uma métrica, a qual denominamos de *activity-level* (AL), para quantificar o interesse do usuário em uma subárea no intuito de satisfazer a essas duas restrições. Calculamos o nível de atividade do usuário u em cada subárea $a_i \in A$ pela razão da frequência f de u em a sobre a soma de todos os *check-ins* de u , como a Equação 1. Diretamente essa informação satisfaz a R2, pois mede a atividade do usuário, e indiretamente a R1, pois utiliza apenas os *check-ins* dos lugares que o usuário foi capaz de visitar.

$$AL_{u,a} = \frac{f_{u,a}}{\sum_{i=1}^N f_{u,a_i}} \quad (1)$$

Sobretudo, esse *activity-level* representa a frequência proporcional de u na subárea a . Subáreas com valores de *activity-level* mais altos representam aquelas onde o usuário realizou muitos *check-ins* durante o período de tempo compreendido pelos

dados de treinamento. Por outro lado, subáreas com valores de *activity-level* iguais ou próximos de 0 representam aquelas onde u não frequenta ou é pouco frequente, respectivamente. Vale destacar que é possível expandir a maneira de calcular esse *activity-level* para considerar outros fatores, tais como relações sociais ou características temporais. Contudo, essas análises não abrangem o escopo deste trabalho.

3.3. Estratégia de Pós-processamento

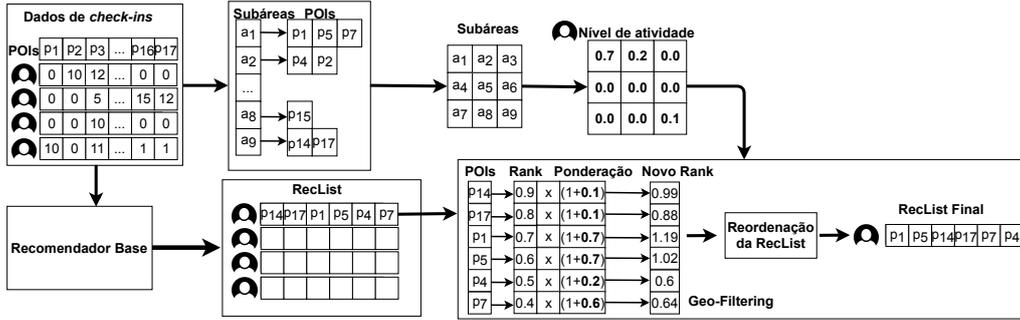


Figura 2. Exemplo da aplicação da etapa de pós-processamento proposta.

Com os valores de *activity-level* calculados para cada usuário, nosso método consiste em uma nova abordagem de pós-processamento para recomendação de POIs. Basicamente, assumimos que existe um recomendador base sendo utilizado no sistema a fim de apresentar uma lista de N possíveis POIs a serem visitados. Esse recomendador estima a relevância r_i de cada POI candidato de acordo com sua metodologia implementada e os reordena de forma decrescente de relevância. A partir dessas informações, executamos nossa estratégia de pós-processamento nessa lista temporária, agregando o valor de relevância desses POIs, originais do recomendador base, com o *activity-level* $AL_{u,a}$. Portanto, para cada POI da lista temporária, calculamos a utilidade $f(u, i)$ de acordo com a Equação 2, que multiplica as notas de relevância retornadas pelo recomendador base pelo valor do *activity-level* do $AL_{u,a}$, onde a consiste na subárea em que o POI está geograficamente localizado. Ao final do pós-processamento, reordenamos decrescentemente a lista temporária de acordo com esses novos valores de relevância e selecionamos os k itens mais relevantes, onde $k \ll N$, como a lista final de recomendações *RecList*.

$$f(u, i) = r_i * (1 + AL_{u,a}) \quad (2)$$

Na Figura 2 apresentamos um exemplo de nossa estratégia. Nesse exemplo, temos que o recomendador base retorna a lista dos seis POIs mais relevantes para um usuário, nessa ordem: $P14$, $P17$, $P1$, $P5$, $P4$ e $P7$. O $P1$ encontra-se na terceira posição e possui uma relevância de 0.7 e está situado na sub-região 1 que representa a subárea de maior atividade do usuário com nível igual a 0.7. Aplicando-se então nossa etapa de pós-processamento ponderando a relevância pela equação acima temos que a nova relevância de $P1$ é dada por: $0.7 * (1 + 0.7) = 1.19$. Após a aplicação desse processo em todos os POIs e reordenando a lista pelos novos valores de relevância, temos que o POI $P1$ passa a ocupar a primeira posição da lista. Esperamos que nosso pós-processamento seja capaz de atribuir mais relevância aos POIs que estão em subáreas onde o usuário mais frequentou no passado. Além disso, o pós-processamento também diminui a relevância final de POIs que estão em subáreas que foram pouco frequentadas pelo usuário, mas que poderiam estar no topo da lista de recomendações temporária fornecida pelo recomendador base.

4. Resultados experimentais

4.1. Configuração Experimental

4.1.1 - Base de Dados: Para realização dos experimentos selecionamos os dados de 3 das 5 cidades que possuem maior número de POIs na base do *Yelp Dataset Challenge*¹: **(1) Charlotte** (2.445 usuários e 7.151 POIs); **(2) Las Vegas** (15.168 usuários e 11.562 POIs); e **(3) Phoenix** (14.388 usuários e 36.827 POIs). Filtramos as informações desse conjunto de dados selecionando os POIs que possuem pelo menos 5 *check-ins* e os usuários que realizaram no mínimo 20 visitas em cada cidade [Han and Yamana 2017]. Em seguida, para cada uma das cidades, reordenamos os *check-ins* pela data em que foram realizados, i.e., do mais antigo para o mais recente. Assim, criamos um conjunto com os 70% primeiros *check-ins* para realizarmos o treinamento dos algoritmos e outro com os 30% restantes para executarmos os testes. Realizamos essa abordagem para simular um cenário de aplicação real do modelo [Zhang and Chow 2015].

4.1.2 - Algoritmos de Recomendação: Para validar nossa proposta, selecionamos estratégias de recomendação propostas em dois contextos distintos. Selecionamos os algoritmos clássicos que estão implementados na biblioteca *MyMediaLite*²: **(1) Most-Popular:** Recomenda uma lista dos POIs mais populares que ainda não foram visitados por ele. A popularidade dos POIs é medida pelo número de visitas que cada POI possui; **(2) User-kNN:** Recomenda os POIs mais relevantes dos K usuários mais similares ao usuário alvo. A similaridade dos usuários é calculada com base na similaridade cosseno e são selecionados pelo menos 80 vizinhos do usuário alvo; e **(3) WRMF:** Recomenda os POIs mais bem avaliados com base na associação dos fatores latentes dos usuário com os dos itens, modelado por um método de Fatoração de Matriz (e.g., SVD, PCA, etc).

Selecionamos outras três abordagens propostas para o cenário de recomendação de POIs: **(1) USG:** Técnica que recomenda POIs levando em consideração a influência geográfica, social e as preferências do usuário. A influência geográfica é dada pela probabilidade de um usuário visitar um POI dado seu histórico de *check-ins* e as distâncias entre POIs. A influência social é usada para definir o perfil do usuário e leva em conta nas preferências dos amigos das redes sociais [Ye et al. 2011]; **(2) GeoMF:** É um modelo de recomendação de fatoração de matrizes ponderado pelas informações geográficas. O fator geográfico é explorado definindo a possibilidade de um usuário visitar uma região, bem como a influência de cada POI em cada região [Lian et al. 2014]. **(3) GeoSoCa:** Técnica de recomendação que explora os fatores geográficos, sociais e categóricos para as recomendações. A influência geográfica é definida por um método de estimativa de *kernel* personalizado pela distribuição de *check-ins* de cada usuário. [Zhang and Chow 2015]

4.1.3 - Métricas de Avaliação: selecionamos as métricas clássicas de *Precision*, *Recall*, *nDCG*, e *MRR* [Zhang and Chow 2015]. Enquanto *Precision* retorna a fração dos POIs relevantes recuperados nas recomendações, *Recall* retorna a fração de POIs relevantes recuperados dentre os POIs visitados pelos usuários. Por sua vez, *nDCG* avalia a qualidade da recomendação baseada na ordem que os POIs relevantes aparecem e *MRR* mensura o esforço do usuário para encontrar o primeiro POI relevante para ele. Calculamos os valores dessas métricas para cada usuário, e depois retornamos a média dos valores.

¹<https://www.yelp.com/dataset/challenge>

²<http://mymedialite.net/>

4.2. Resultados

4.2.1 - Recomendadores Tradicionais

Primeiramente, selecionamos os três SsR tradicionais, chamados de *MostPopular*, *UserKNN* e *WRMF* e os aplicamos nas bases de dados relacionadas à cidades de Las Vegas, Phoenix e Charlotte, quantificando as métricas de avaliação selecionadas. Em seguida, a partir de recomendações geradas por cada um dos algoritmos, aplicamos nosso processo de pós-processamento, novamente medindo as mesmas métricas. Ambos os valores são sumarizados na Tabela 1. Realizamos uma avaliação estatística, utilizando o teste de Wilcoxon [Wilcoxon 1945] com 95% de intervalo de confiança. O símbolo ▲ representa ganhos positivos estatisticamente significativos, ● representa empates estatísticos e ▼ representa perdas estatisticamente significativas.

	Recommender	Precision		Recall		nDCG		MRR	
		top-10	top-15	top-10	top-15	top-10	top-15	top-10	top-15
Las Vegas	<i>Most Popular</i>	0.0153	0.0155	0.0266	0.0395	0.0205	0.0260	0.0450	0.0491
	<i>Geo-Filtering</i>	0.0172 ▲	0.0167 ▲	0.0331 ▲	0.0459 ▲	0.0248 ▲	0.0304 ▲	0.0527 ▲	0.0569 ▲
	<i>User-kNN</i>	0.0311	0.0279	0.0583	0.0755	0.0476	0.0552	0.0971	0.1020
	<i>Geo-Filt</i>	0.0317 ▲	0.0285 ▲	0.0606 ▲	0.0777 ▲	0.0489 ▲	0.0565 ▲	0.0994 ▲	0.1043 ▲
	<i>WRMF</i>	0.0327	0.0296	0.0592	0.0767	0.0485	0.0567	0.0994	0.1043
Phoenix	<i>Geo-Filtering</i>	0.0332 ▲	0.0301 ▲	0.0613 ▲	0.0794 ▲	0.0500 ▲	0.0582 ▲	0.1027 ▲	0.1077 ▲
	<i>Most Popular</i>	0.0129	0.0118	0.0089	0.0119	0.0096	0.0115	0.0356	0.0379
	<i>Geo-Filtering</i>	0.0131 ▲	0.0123 ▲	0.0090 ▲	0.0127 ▲	0.0098 ▲	0.0120 ▲	0.0361 ▲	0.0387 ▲
	<i>User-kNN</i>	0.0327	0.0294	0.0244	0.0329	0.0276	0.0328	0.0973	0.1022
	<i>Geo-Filtering</i>	0.0329 ●	0.0298 ▲	0.0247 ▲	0.0336 ▲	0.0279 ●	0.0332 ●	0.0979 ●	0.1030 ●
Charlotte	<i>WRMF</i>	0.0276	0.0255	0.0204	0.0278	0.0226	0.0272	0.0797	0.0845
	<i>Geo-Filtering</i>	0.0281 ▲	0.0259 ▲	0.0208 ▲	0.0284 ▲	0.0231 ▲	0.0277 ▲	0.0812 ▲	0.0858 ▲
	<i>Most Popular</i>	0.0367	0.0338	0.0280	0.0386	0.0290	0.0352	0.0971	0.1024
	<i>Geo-Filtering</i>	0.0375 ▲	0.0344 ▲	0.0285 ●	0.0400 ▲	0.0297 ●	0.0361 ●	0.0978 ●	0.1033 ●
	<i>User-kNN</i>	0.0481	0.0441	0.0399	0.0544	0.0416	0.0500	0.1317	0.1378
Charlotte	<i>Geo-Filtering</i>	0.0478 ▼	0.0443 ●	0.03963 ●	0.0547 ●	0.0416 ●	0.0503 ●	0.1325 ●	0.1390 ●
	<i>WRMF</i>	0.0492	0.0444	0.0393	0.0525	0.0424	0.0503	0.1375	0.1433
Charlotte	<i>Geo-Filtering</i>	0.0499 ●	0.0450 ●	0.0404 ●	0.0539 ▲	0.0432 ●	0.0512 ●	0.1392 ●	0.1450 ●

Tabela 1. Resultados dos recomendadores tradicionais sem utilizar e utilizando a estratégia de pós-processamento.

É importante destacar que dentre esses métodos, nenhum explora a influência geográfica como um fator decisivo nas recomendações. Portanto, ao dividirmos uma região em sub-regiões, aplicando nossa proposta, a filtragem geográfica consegue computar uma informação que antes não era explorada. Com isso, podemos observar uma melhora significativa dos resultados. Na cidade de Las Vegas encontramos ganhos estatisticamente significativos quando comparados com a lista original. O mesmo acontece quando comparamos os resultados na cidade de Phoenix. Especificamente, na técnica *MostPopular* os resultados são mais explícitos, uma vez que a filtragem geográfica reordena a lista de forma personalizada, visando atender melhor o padrão de consumo dos usuários.

Na cidade de Charlotte, os resultados não são tão expressivos. Na grande maioria das combinações, nossa etapa de pós-processamento apresenta empate estatísticos com os resultados originais dos algoritmos. Isso acontece pois essa é a cidade com menor número de usuários e uma pequena quantidade de POIs por subárea. Além disso, o número de POIs visitados por cada usuário também é baixo. Nesse sentido, a pouca informação de consumo dos usuários afeta consideravelmente a qualidade dos recomendadores tradicionais. Como consequência, a lista de POIs recomendados tende a conter itens pouco relevantes para os usuários. Nossa estratégia é puramente dependente da lista de recomendação e da área de atuação de cada usuário. Dessa forma, reordenar uma lista de POIs que, em sua maioria, já não apresenta itens relevantes para os usuários, pouco interfere na qualidade final.

4.2.2 - Recomendadores de POIs

Em nossa segunda análise, selecionamos três SsR específicos para POIs como linhas de base que são consideradas estado-da-arte nesse domínio: *USG* [Ye et al. 2011], *GeoMF* [Lian et al. 2014] e *GeoSoCa* [Zhang and Chow 2015]. Novamente, recuperamos a lista de recomendação gerada e aplicamos a nossa proposta de filtragem geográfica. Todos os resultados obtidos podem ser observados na Tabela 2.

		Precision		Recall		nDCG		MRR	
Recommender		top-10	top-15	top-10	top-15	top-10	top-15	top-10	top-15
Las Vegas	<i>USG</i>	0.0276	0.0257	0.0545	0.0721	0.0442	0.0520	0.0889	0.0939
	<i>Geo-Filtering</i>	0.0290 ▲	0.0266 ▲	0.0583 ▲	0.0759 ▲	0.0466 ▲	0.0544 ▲	0.0926 ▲	0.0975 ▲
	<i>GeoMF</i>	0.0284	0.0255	0.0546	0.0713	0.0437	0.0510	0.0896	0.0942
	<i>Geo-Filtering</i>	0.0291 ▲	0.0261 ▲	0.0551 ●	0.0732 ▲	0.0446 ▲	0.0522 ▲	0.0923 ▲	0.0971 ▲
	<i>GeoSoCa</i>	0.0206	0.0180	0.0339	0.0439	0.0282	0.0326	0.0660	0.0691
	<i>Geo-Filtering</i>	0.0201 ▼	0.0177 ●	0.0322 ▼	0.0430 ●	0.0271 ▼	0.0318 ▼	0.0628 ▼	0.0661 ▼
Phoenix	<i>USG</i>	0.0228	0.0212	0.0176	0.0241	0.0198	0.0238	0.0701	0.0744
	<i>Geo-Filtering</i>	0.0231 ▲	0.0215 ▲	0.0179 ▲	0.0246 ▲	0.0202 ▲	0.0243 ▲	0.0714 ▲	0.0758 ▲
	<i>GeoMF</i>	0.0297	0.0278	0.0235	0.0330	0.0256	0.0312	0.0874	0.0928
	<i>Geo-Filtering</i>	0.0302 ▲	0.0281 ▲	0.0242 ▲	0.0336 ▲	0.0263 ●	0.0318 ●	0.0893 ●	0.0946 ●
	<i>GeoSoCa</i>	0.0171	0.0154	0.0112	0.0151	0.0131	0.0156	0.0516	0.0543
	<i>Geo-Filtering</i>	0.0171 ●	0.0155 ●	0.0111 ●	0.0153 ●	0.0131 ●	0.0157 ●	0.0512 ●	0.0539 ●
Charlotte	<i>USG</i>	0.0420	0.0387	0.0334	0.0457	0.0338	0.0409	0.1088	0.1143
	<i>Geo-Filtering</i>	0.0432 ▲	0.0394 ▲	0.0347 ▲	0.0475 ▲	0.0350 ▲	0.0423 ▲	0.1111 ●	0.1172 ▲
	<i>GeoMF</i>	0.0349	0.0325	0.0282	0.0391	0.0300	0.0364	0.0992	0.1050
	<i>Geo-Filtering</i>	0.0357 ●	0.0334 ▲	0.0299 ▲	0.0409 ▲	0.0313 ●	0.0377 ▲	0.1028 ▲	0.1084 ▲
	<i>GeoSoCa</i>	0.0230	0.0218	0.0163	0.0235	0.0183	0.0226	0.0630	0.0674
	<i>Geo-Filtering</i>	0.0237 ●	0.0215 ●	0.0167 ●	0.0232 ●	0.0189 ●	0.0227 ▲	0.0636 ▲	0.0676 ▲

Tabela 2. Resultados recomendadores específicos para POIs sem utilizar e utilizando a estratégia de pós-processamento.

Para esse caso, nosso modelo atinge bons resultados para as técnicas de recomendação *USG* e *GeoMF* nas três cidades selecionadas. Isso se deve ao fato de que nossa proposta consegue auxiliar os SsR específicos de POIs, que lidam com o fator geográfico. Basicamente, nesses casos, o nível de atividade definida pela proposta desse trabalho agrega as informações geográficas nas recomendações, levando em conta a densidade de atividade de cada usuário em sub-regiões. A diferença mais acentuada pode ser notada no pós-processamento do *USG*, em Las Vegas e Charlotte. Novamente, obtivemos ganhos estatisticamente significativos (Wilcoxon com 95% de confiança). Tais resultados reforçam a ideia de que nossa proposta atinge os objetivos e contribuições esperadas, especialmente, sobre o modelo do *GeoMF*, um dos principais da literatura [Liu et al. 2017].

Ao comparar nossa proposta com o *GeoSoCa* [Zhang and Chow 2015] o desempenho do pós-processamento diminui, tanto em Las Vegas quanto em Phoenix. Isso acontece pois o *GeoSoCa* agrega dois atributos a mais em suas recomendações, a influência social e categórica. Além disso, segundo *Liu et. al* [Liu et al. 2017], o fator social explorado pelo *GeoSoCa* é o que possui melhor desempenho se comparado com outros algoritmos (e.g., o *USG*). Portanto, nossa proposta trata somente com as características geográficas de cada usuário, individualmente, justificando assim essa diferença encontrada. Mesmo sendo inferior, os resultados obtidos foram muito próximos dos encontrados pela linha de base.

Um resultado interessante pode ser observado na cidade de Charlotte na qual nossa estratégia consegue melhorar os resultados originais do *GeoSoCa*. Isso ocorre porque Charlotte possui menos dados de correlação social, onde cada usuário, em média, possui apenas um amigo. Logo, a agregação de informação que o fator social traz para o *GeoSoCa* não é relevante. Assim, a aplicação de nossa etapa de pós-processamento é capaz de enriquecer a informação geográfica trazendo melhorias à recomendação do *GeoSoCa*.

5. Conclusões e Trabalhos Futuros

Neste trabalho propomos um modelo de pós-processamento que pondera e filtra um conjunto de recomendações geradas por um recomendador base a fim de satisfazer ainda mais as preferências dos usuários. Para tal, construímos uma métrica capaz de quantificar o nível de atividade do usuário em sub-regiões. Nossa premissa é que ao satisfazer as duas restrições impostas pela lei da geografia de Tobler, favorecemos ainda mais a satisfação dos usuários. De maneira geral, observamos que a inclusão do *Geo-filtering*, como pós-processamento, permite de maneira simples e eficaz aumentar a acurácia das recomendações. Essa nova etapa de pós-processamento acrescenta características ainda não exploradas por outras linhas de base, mostrando que, dividir uma área em subáreas menores, calculando para cada usuário seu nível de atividade, possibilita extrair as sub-regiões mais atuantes dos usuários. Os resultados encontrados mostram ganhos estatisticamente significativos na maioria dos casos, principalmente em grandes cidades, como Phoenix e Las Vegas.

Como mencionado anteriormente, nossa proposta ainda está limitada a linhas de base que exploram outras características além da geográfica, como, por exemplo, o *GeoSoca*. Futuramente, nosso objetivo é incluir também outras características, como a influência social no cálculo do nível de atividade do usuário a fim de extrair mais informações não avaliadas pelos SsR. Também pretendemos incluir outras formas de se obter as sub-regiões, dividindo as subáreas de forma variada ou incluindo novas subáreas dentro das já divididas.

Referências

- A., G., Ramos, G., Madeira, D., Sachetto, R., Ferreira, R., and Rocha, L. (2013). G-dbscan: A gpu accelerated algorithm for density-based clustering. *Procedia Computer Science -ICCS*, 18(0):369 – 378.
- Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46:109–132.
- Fernández-Tobías, I., Tomeo, P., Cantador, I., Di Noia, T., and Di Sciascio, E. (2016). Accuracy and diversity in cross-domain recommendations for cold-start users with positive-only feedback. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 119–122. ACM.
- Gao, H., Tang, J., Hu, X., and Liu, H. (2013). Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 93–100. ACM.
- Han, J. and Yamana, H. (2017). Geographical diversification in poi recommendation: toward improved coverage on interested areas. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 224–228. ACM.
- Kumar, S. S., Ramachandran, T., and Panboli, S. (2015). Product recommendations over facebook: the roles of influencing factors to induce online shopping. *Asian Social Science*, 11(2):202.
- Li, X., Cong, G., Li, X.-L., Pham, T.-A. N., and Krishnaswamy, S. (2015). Rank-geofm: A ranking based geographical factorization method for point of interest recommenda-

- tion. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 433–442. ACM.
- Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., and Rui, Y. (2014). Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 831–840. ACM.
- Liu, B., Fu, Y., Yao, Z., and Xiong, H. (2013). Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1043–1051. ACM.
- Liu, Y., Pham, T.-A. N., Cong, G., and Yuan, Q. (2017). An experimental evaluation of point-of-interest recommendation in location-based social networks. *Proceedings of the VLDB Endowment*, 10(10):1010–1021.
- Liu, Y., Wei, W., Sun, A., and Miao, C. (2014). Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 739–748. ACM.
- Lu, J., Wu, D., Mao, M., Wang, W., and Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74:12–32.
- Miller, H. J. (2004). Tobler’s first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2):284–289.
- Silva, N., Carvalho, D., Pereira, A. C., Mourão, F., and Rocha, L. (2019). The pure cold-start problem: A deep study about how to conquer first-time users in recommendations domains. *Information Systems*, 80:1–12.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.
- Ye, M., Yin, P., and Lee, W.-C. (2010). Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 458–461. ACM.
- Ye, M., Yin, P., Lee, W.-C., and Lee, D.-L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR*, pages 325–334. ACM.
- Yu, Y. and Chen, X. (2015). A survey of point-of-interest recommendation in location-based social networks. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Yuan, Q., Cong, G., Ma, Z., Sun, A., and Thalmann, N. M. (2013). Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR*, pages 363–372. ACM.
- Zhang, J.-D. and Chow, C.-Y. (2015). Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *Proceedings of the 38th International ACM SIGIR*, pages 443–452. ACM.