

Análisis de estrategias de selección de vecinos para recomendación en LBSN

Carlos Rios, Daniela Godoy y Silvia Schiaffino

ISISTAN, UNCPBA. CONICET. Facultad de Ciencias Exactas. Campus Universitario, Tandil (B7001BBO), Buenos Aires, Argentina. Tel.: +54 (249) 4439682 ext. 35. Fax.: +54 (249) 4439681 carlos.rios@isistan.unicen.edu.ar

Abstract. El gran uso de los dispositivos móviles y los servicios basados en ubicación han generado un nuevo concepto en los medios sociales en línea, llamado redes sociales basadas en ubicación. Éstas usan tecnologías como GPS, Web 2.0 y smartphones para permitir a los usuarios compartir sus ubicaciones (check-ins), buscar lugares de interés o POIs (Point of Interest), descuentos, dejar comentarios de lugares específicos, conectarse con sus amigos y encontrar amigos que se encuentran cerca de algún lugar específico. Para aprovechar la información que los usuarios vuelcan en estas redes surgieron los Sistemas de Recomendación basados en Ubicación (LBSNs, sus siglas en inglés) que generan sugerencias en base a la aplicación de diferentes técnicas de recomendación. En este artículo se presentan dos estrategias para la selección de vecinos en el enfoque de filtrado colaborativo clásico basado en usuarios, considerando la red social de los usuarios y las visitas comunes como factores influyentes. El enfoque propuesto fue evaluado utilizando datos de una red social basada en ubicación popular, mostrando mejoras sobre el enfoque clásico de filtrado colaborativo.

1 Introducción

La gran explosión en el uso de los celulares, la facilidad para adquirir la localización geográfica y el desarrollo de las comunicaciones wireless, ha permitido la creación de servicios sociales que tienen como principal característica la ubicación geográfica del usuario. Foursquare¹ es la red social más popular dentro de este tipo de servicios, y permite a los usuarios fácilmente compartir su ubicación geográfica y los contenidos relacionados a dicha ubicación en forma online. La ubicación geográfica es una nueva dimensión en las redes sociales que disminuye la brecha entre el mundo físico y los servicios de redes sociales en línea, dando lugar a nuevas oportunidades y desafíos a los sistemas de recomendación tradicionales. Los sistemas de recomendación son una manera de atacar el problema de sobrecarga de información que se enfrenta el usuario mientras busca información acerca de elementos o ítems de interés. Los métodos de recomendación tradicionales como Filtrado Colaborativo (FC), Basado en Contenido (BC) e híbridos [9] procesan la información derivada de las calificaciones brindadas por los usuarios y de las características de los ítems para realizar sus recomendaciones. Sin embargo, en las redes sociales existe información adicional que los métodos de recomendación deben tener en cuenta, tales como el comportamiento del usuario y las

¹ Foursquare Web Site, <https://foursquare.com>

relaciones de amistad entre ellos [13]. Con respecto a las redes sociales basadas en la ubicación (LBSNs), los datos geo-localizados son una dimensión física que las redes sociales tradicionales no poseen. En esta nueva era, los usuarios pueden beneficiarse de obtener un acceso omnipresente y ubicuo a los servicios basados en ubicación desde cualquier lugar a través de dispositivos móviles. Las distintas fuentes de datos generadas a partir de una red social basada en la ubicación, generan relaciones de diversos tipos: la relación Usuario-Usuario, mostrando la amistad entre ellos o la coincidencia en lugares visitados; la relación Usuario-Lugar, mostrando que un usuario visitó dicho lugar; la relación Lugar-Lugar, la cual muestra relación de categorías o relación de distancia. Además de estas relaciones los usuarios generan contenido, como por ejemplo comentarios luego de visitar un lugar. En este contexto, surgen los Sistemas de Recomendación basados en Ubicación [11], que usan la propiedad geográfica como una fuente auxiliar para realizar recomendaciones de amigos [7,10], lugares [12,6], actividades [3,14] y eventos [5,8]. La heterogeneidad de los datos producidos por las redes sociales basadas en ubicación genera la necesidad de nuevos enfoques dentro de los sistemas de recomendación, utilizando diferentes fuentes de datos y metodologías para realizar distintos tipos de recomendación.

En este trabajo se proponen dos estrategias para la selección de vecinos en un marco del método de Filtrado Colaborativo basado en usuario para la recomendación de lugares, utilizando las relaciones generadas en una red social basada en ubicación. La hipótesis principal de este trabajo es que la limitación de los vecinos potenciales puede reducir la comparación en la búsqueda en la selección de vecinos y además mejorar la estimación de la preferencia.

El resto de este artículo está organizado como sigue: la Sección 2 discute trabajos relacionados en el tema. La Sección 3 presenta el enfoque propuesto. La Sección 4 muestra los resultados experimentales realizados con un conjunto de datos reales tomados de la red social Foursquare y, finalmente, las conclusiones alcanzadas con este trabajo se detallan en la Sección 5.

2 Trabajos relacionados

Los sistemas de recomendación en LBSNs usan la propiedad geográfica como una fuente auxiliar para mejorar sus recomendaciones de lugares o actividades en el que un usuario objetivo puede estar interesado. Generalmente, los sistemas de recomendación en LBSNs se basan en los algoritmos tradicionales de recomendación. La principal diferencia es que en los sistemas de recomendación en LBSNs la entrada elemental de los algoritmos de recomendación es el historial de ubicaciones del usuario. Los distintos tipos de recomendación pueden ser, amigos, lugares (POI), actividades y eventos.

Estudios recientes reflejan la importancia de la ubicación en la generación de nuevos vínculos de amistad. Por ejemplo, en [10] se estudió el problema de recomendación de amigos y se concluyó que cerca del 30% de los nuevos enlaces son elegidos entre usuarios que visitaron los mismos lugares. En [4] los autores proponen un enfoque donde el principal concepto es recomendar amigos que tienen similar interés u otras cosas similares con el usuario objetivo. Utilizan la información de las redes sociales tales como intereses, el concepto de la ubicación en la vida real y el tiempo de permanencia. Cuando

un usuario está en un lugar desconocido, probablemente quiere saber dónde dirigirse por ejemplo en busca de comida o para hacer turismo. Las recomendaciones de lugares o POIs pueden proporcionar una guía en este sentido. En [12] sostienen que la influencia geográfica entre los POIs juega un importante rol en el comportamiento de check-ins de los usuarios y lo modelan como una distribución “power law”. Además, proponen un framework de recomendación de POI, el cual fusiona las preferencias del usuario a POI con influencia social y geográfica. Una actividad es consistente con la ubicación donde el usuario realiza el check-in. En [3] los autores proponen un enfoque para detectar el contexto actual del usuario, inferir posibles actividades de ocio y recomendar contenidos adecuados sobre el lugar (tiendas, parques, películas). La recomendación de eventos es un caso especial de recomendación de actividades, donde se recomienda eventos sociales (conciertos, fiestas, encuentros, etc) que se llevaran a cabo en un lugar específico. A diferencia de estos trabajos en este artículo proponemos un análisis de uno de los pasos influyente en el metodo de recomendación filtrado colaborativo basada en usuarios, el cual es la selección de vecinos.

3 Enfoques de selección de vecinos

En un escenario tradicional de FC, hay una lista de m usuarios $U = u_1, u_2, \dots, u_m$ y una lista de n ítems $I = i_1, i_2, \dots, i_n$ que pueden ser recomendados a los usuarios. Cada usuario u_i ha expresado su opinión acerca de una serie de ítems $Iu_i \in I$, usualmente de manera explícita como una puntuación o calificación en una escala numérica determinada. Esta información se expone en un matriz M de usuarios-ítems de tamaño $m \times n$, de tal manera que el valor de cada celda en M representa la puntuación de preferencia (calificación) dada por el usuario i al ítem j .

Los enfoques de FC basados en memoria hacen predicciones recorriendo la matriz de dos formas, en base a usuarios o en base a ítems [1]. Dado un usuario activo que requiere una predicción para un ítem sin calificación, los algoritmos de FC miden las similitudes entre los usuarios activos y otros usuarios (enfoque basado en el usuario) o entre el ítem y el resto de los ítems (enfoque basado en ítems). Por lo tanto, la calificación se predice mediante la agregación de las calificaciones que el ítem recibió de usuarios similares en el primer caso, o las calificaciones otorgadas por el usuario activo a elementos similares en el segundo enfoque.

El modelo de FC basado en usuarios clásico se define entonces como la ecuación 1:

$$\tilde{r}(u, i) = \bar{r}(u) + C_o \sum_{v \in N_k(u, i)} sim(u, v)(r(v, i) - \bar{r}(v)) \quad (1)$$

donde $r(v, i)$ es la calificación que dio el usuario v al ítem i , \tilde{r} es la predicción de la calificación (a diferencia del rating o calificación observado r), $N_k(u, i)$ es el conjunto de los k usuarios más similares a u y $sim(u, v)$ es la función que mide la similitud entre dos usuarios u y v . C_o es un factor de normalización. La preferencia del usuario u para un ítem i se predice en base a la calificación media $\bar{r}(u)$, la suma de las desviaciones de las calificaciones que dieron todos los vecinos v al ítem i y las calificaciones promedio $\bar{r}(v)$, ponderado por las similitudes con sus vecinos.

Los enfoques basados en usuarios asumen que no todos los usuarios son igualmente útiles en la predicción destinada a un usuario dado, con lo que se desprenden dos problemas principales: (1) cómo seleccionar los vecinos de un usuario para generar recomendaciones; (2) cómo utilizar adecuadamente la información provista por dichos vecinos en la recomendación. Normalmente, la identificación de vecinos está basada en su similitud con el usuario activo, mientras que una práctica común es definir un número máximo de usuarios para acotar el vecindario. Una vez delimitado el vecindario, el aporte de cada vecino a la predicción se pesa o pondera en base a su distancia con el usuario activo. Por ejemplo, una alternativa muy usada es la combinación lineal de las calificaciones de los vecinos pesados por similitud. Existen sin embargo factores adicionales a la similitud que pueden ser valiosos para la selección de vecinos, por ejemplo en el caso de este trabajo el historial de visitas de los usuarios puede considerarse de relevancia más allá de la similitud en las calificaciones.

Con respecto a la selección de vecinos, que plantea la búsqueda de usuarios similares al usuario objetivo, en un enfoque clásico esta tarea es realizada comparando al usuario objetivo con todos los usuarios de la base de datos, produciendo una gran cantidad de comparaciones, y además permitiendo que cualquier usuario similar al usuario objetivo pueda contribuir a la estimación de la preferencia. Es por ello que utilizar las relaciones producidas en las redes sociales basadas en la ubicación para restringir por alguna característica los vecinos potenciales de un usuario objetivo puede reducir el número de comparaciones y tal vez mejorar la estimación de la preferencia. Los siguientes son enfoques que proponemos para la selección de vecinos potenciales explotando la información generada en las redes sociales basadas en la ubicación.

El primer enfoque es utilizar la información de la red social de los usuarios. Podemos limitar el conjunto de vecinos que pueden contribuir a la estimación de la preferencia al grupo de usuarios que se relacionan socialmente con el usuario objetivo. Estas pueden ser relaciones directas (amigos directos) o relaciones indirectas (amigos de mis amigos). En otras palabras se puede buscar los k usuarios relacionados socialmente más similares del usuario objetivo para predecir las preferencias.

Además de la relación social mencionada anteriormente, los usuarios pueden estar relacionados por los lugares comunes visitados entre ellos. Estos pueden tomar una representación de red, donde los usuarios son los nodos y las aristas representan el número de veces que coincidieron en un lugar, produciendo una relación geo localizada. Con la misma idea planteada en la red social, en un segundo enfoque, se puede limitar el conjunto de usuarios que pueden ser potenciales vecinos del usuario objetivo. Por lo tanto se busca los k usuarios más similares del conjunto de usuarios con relación geo localizada con el usuario objetivo.

4 Evaluación experimental

En esta sección se describen los experimentos realizados y los resultados obtenidos, como así también el procesamiento del conjunto de datos utilizado.

4.1 Diseño de los experimentos

Con el objetivo de evaluar el enfoque propuesto se diseñaron experimentos que mostraron el desempeño de las estrategias para la selección de vecinos. Todos los experimentos utilizaron distintos tamaños de vecindario: 5, 10, 20, 30, 50. Se utilizó una función coseno para buscar los k más similares. Luego los experimentos quedaron configurados de la siguiente forma:

- Experimento 1: Vecinos potenciales extraídos de la red social. Se utilizaron distintos niveles de relación de amistad, directos nivel 1 e indirectos nivel 2, 3, 4 y 5.
- Experimento 2: Vecinos potenciales extraídos de la red de visitas. Se utilizaron distintos niveles de relación de geolocalización, directos nivel 1 e indirectos nivel 2, 3, 4 y 5.
- Experimento 3: Línea base de comparación. Se utilizó la selección de vecinos clásica para tener un punto de comparación.

Para los distintos experimentos se realizó un procesamiento de datos que son detallados en la siguiente sección del informe. Para la comparación de las estrategias se utilizó una medida muy utilizada en la bibliografía, MAE (Mean Absolute Error) dada por la ecuación 2.

$$MAE = \frac{1}{|T|} \sum_{(u,i) \in T} |\bar{r}_{ui} - r_{ui}| \quad (2)$$

donde \bar{r}_{ui} es la estimación de la preferencia del ítem i para el usuario u y r_{ui} es la preferencia real del ítem i dada por el usuario u . La métrica mide el error en la estimación de la preferencia, en el conjunto T de pares usuario-item.

4.2 Procesamiento de datos

Para realizar los experimentos planteados se utilizó un conjunto de datos extraído de la red geolocalizada Foursquare, [2].

El conjunto de datos contiene información de: Lugares, archivo que contiene información de los lugares visitados. Usuarios, archivo que contiene información de los usuarios. Tips, archivo que contiene información de los check-in's realizados por los usuarios. Amistad, archivo que contiene información de la relación social entre los usuarios y Categorías, archivo json que contiene información de las categorías de los lugares en Foursquare.

En la base de datos existen usuarios de todo el mundo pero para realizar los experimentos se filtraron los usuarios que pertenecen al estado de New York, ya que son mayor en cantidad. Por lo tanto la distribución quedó con 47220 usuarios totales, de los cuales 27000 son usuarios de NY. Para el experimento 1, se utilizó la red de amistad de Foursquare, la cual posee las siguientes características: red no dirigida, sin peso en las aristas, con un total de 47220 nodos y 1192758 aristas.

Para el experimento 2, se procesaron los datos para generar la red de visitas comunes. En esta red un nodo es un usuario y la relación entre ellos esta dada por la

cantidad de visitas comunes, se eliminaron las relaciones con 1 (una) visita en común entre los usuarios debido a que estas pueden ser por pura coincidencia. La red quedó categorizada como una red no dirigida, con peso en las aristas(cantidad de visitas comunes entre usuarios), con un total de 47220 nodos y 147086 aristas.

Las visitas fueron utilizadas como una forma de preferencia, ya que si una persona visita muchas veces un lugar se puede deducir que le gustó mucho. En este contexto para la generación de la matriz de preferencias se utilizó una escala de valores de 5 puntos, donde 1 representa “Malo”, 2 “Regular”, 3 “Bueno”, 4 “Muy Bueno”, 5 “Excelente” . Estos se mapearon a la escala de valores de la siguiente manera, 1 visita se mapea con el valor 2, 2 visitas con el valor 3, 3 visitas con el valor 4 y 4 o más visitas con el valor 5.

4.3 Resultados

Los resultados de los experimentos realizados se evaluaron utilizando la medida clásica MAE. Además se contabilizó el número de vecinos potenciales. Los resultados obtenidos mostraron en algunos casos menores errores a la línea base y también un menor número de vecinos potenciales.

La figura 1 arriba muestra los resultados del experimento 1, donde se observa que los menores errores de estimación se dan donde la selección de los K vecinos está dada en el conjunto de vecinos potenciales obtenidos de las relaciones sociales a un nivel 2. Estos representan amigos directos y los amigos de ellos.

Con respecto a la figura 1 abajo, ésta muestra los resultados del experimento 2, donde se observa que los menores errores de estimación se dan donde la selección de los K vecinos está dada en el conjunto de vecinos potenciales obtenidos de la red de visitas a un nivel 1. Estos representan usuarios que coincidieron más de una vez, en los lugares visitados.

Con respecto al número de comparaciones para la selección de vecinos, la figura 2, muestra el promedio de vecinos potenciales en las diferentes estrategias. Como es de esperar el enfoque clásico o base, realiza un escaneo en toda la base de datos, luego la red social nivel 5, 4 y 3, también produce muchos vecinos potenciales, ya que es una acumulación de amigos, amigos de mis amigos, etc. Con respecto a la red de visitas, al eliminar las relaciones de una visita en común la red quedó menos densa generando una reducción de vecinos potenciales en todos sus niveles.

5 Conclusiones y trabajo futuro

El presente trabajo planteó estrategias para la selección de vecinos en un marco del método Filtrado Colaborativo basado en usuario para la recomendación de lugares de interés (POI), se diseñaron experimentos y procesaron datos para evaluar dichas estrategias, los resultados mostraron que utilizar información de distintas fuentes para la selección de vecinos mejora la estimación de las preferencias y reduce considerablemente el número de vecinos potenciales. Con respecto a los trabajos futuros se plantea el proceso de todas las relaciones generadas en una red social geo localizada y utilizarlas tanto como para la selección de vecinos y la ponderación de los mismos.

Fig. 1. Resultados

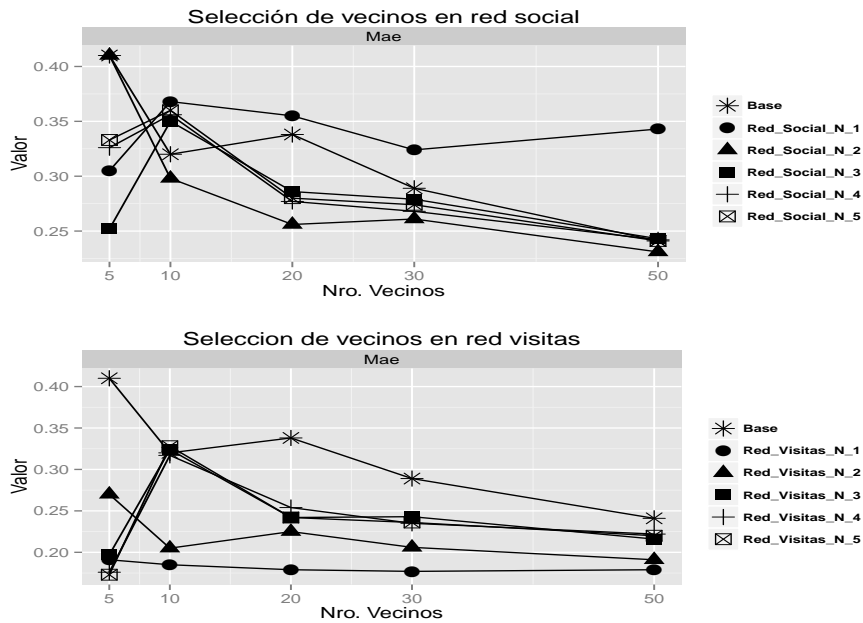
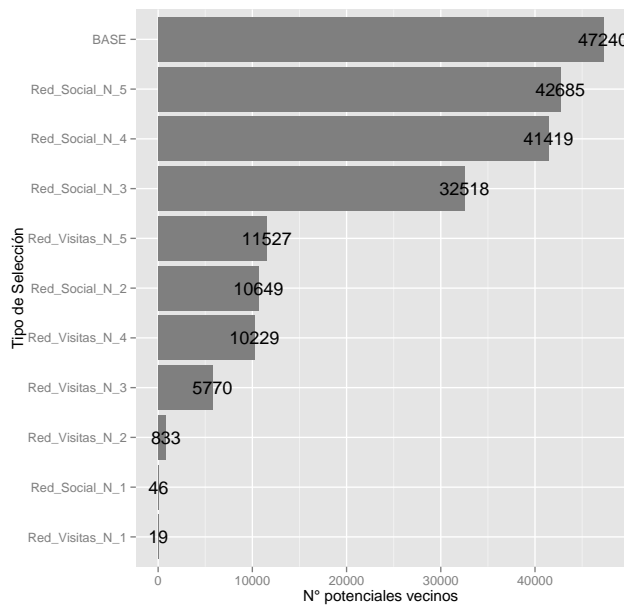


Fig. 2. Cantidad promedio de vecinos



Referencias

1. Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, June 2005.
2. Jie Bao, Yu Zheng, and Mohamed F. Mokbel. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems, SIGSPATIAL '12*, pages 199–208, New York, NY, USA, 2012. ACM.
3. Victoria Bellotti, Bo Begole, Ed H. Chi, Nicolas Ducheneaut, Ji Fang, Ellen Isaacs, Tracy King, Mark W. Newman, Kurt Partridge, Bob Price, Paul Rasmussen, Michael Roberts, Diane J. Schiano, and Alan Walendowski. Activity-based serendipitous recommendations with the magitti mobile leisure guide. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08*, pages 1157–1166, New York, NY, USA, 2008. ACM.
4. Cheng-Hao Chu, Wan-Chuen Wu, Cheng-Chi Wang, Tzung-Shi Chen, and Jen-Jee Chen. Friend recommendation for location-based mobile social networks. In *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2013 Seventh International Conference on*, pages 365–370, July 2013.
5. Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya. Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web*, 14(4):321–349, 2011.
6. Xuelian Long and J. Joshi. A hits-based poi recommendation algorithm for location-based social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, pages 642–647, Aug 2013.
7. Alexis Papadimitriou, Panagiotis Symeonidis, and Yannis Manolopoulos. Friendlink: link prediction in social networks via bounded local path traversal. In *Computational Aspects of Social Networks (CASoN), 2011 International Conference on*, pages 66–71. IEEE, 2011.
8. D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft. Recommending social events from mobile phone location data. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 971–976, Dec 2010.
9. Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 1–35. Springer US, 2011.
10. Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. Exploiting place features in link prediction on location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1046–1054. ACM, 2011.
11. Yannis Symeonidis Panagiotis, Ntempos Dimitrios. *Recommender Systems for Location-based Social Networks*. Springer, 2014.
12. Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11*, pages 325–334, New York, NY, USA, 2011. ACM.
13. R. Zafarani, M.A. Abbasi, and H. Liu. *Social Media Mining: An Introduction*. Cambridge University Press, 2014.
14. A. Zanda, E. Menasalvas, and S. Eibe. A social network activity recommender system for ubiquitous devices. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, pages 493–497, Nov 2011.