

Tandil, 12 de julio de 2013

Sra. Directora del
Dpto. de Computación y Sistemas
Fac. Ciencias Exactas, UNCPBA
Dra. Mariana del Fresno
S / D

De nuestra consideración:

Me dirijo a Ud. con el propósito de solicitar la evaluación del plan de Trabajo Final "*Inferir vínculos sociales entre usuarios con historial de recorridos*", un estudio que tiene como fin implementar y aplicar un algoritmo de evaluación de similitud entre usuarios basándose en sus recorridos, para luego analizar los resultados obtenidos. El mismo será desarrollado bajo la dirección del Mg. Oscar Nigro.

Sin otro particular, lo saluda atte.

Pamela Nibaldi
Legajo 245950

Propuesta de Trabajo Final Carrera de Ingeniería de Sistemas

“Inferir vínculos sociales entre usuarios con historial de recorridos”

Alumna:

Pamela Nibaldi

Director:

Mg. Oscar Nigro

Motivación

El motivo por el cual surge este trabajo de grado se debe al interés de analizar cómo los recorridos que realizan las personas pueden ayudar a inferir similitudes o relaciones entre las mismas.

El proyecto consiste en analizar recorridos semánticos de diversas personas dentro de la ciudad de Tandil, en diferentes días y a lo largo de todo el día. Dicha información provendrá de encuestas realizadas en distintos hogares de la ciudad. Tomando como base el estudio realizado por la Universidad de Hong Kong [1], donde se analiza cómo es posible obtener vínculos entre las personas en función de su historial de movimientos, se va a implementar el algoritmo propuesto con algunas modificaciones, para luego aplicarlo a la información recopilada. Para finalizar, se pretenden analizar los resultados obtenidos.

Marco teórico

Las redes sociales basadas en las ubicaciones se han vuelto populares en los años recientes. En este proyecto, se desea estimar la similitud entre usuarios de acuerdo con sus historias de recorridos. Esta similitud puede ser un vínculo potencial social entre usuarios, permitiendo así realizar recomendaciones de amistad y lugares.

Objetivos

En este proyecto se pretende implementar un algoritmo que en base a dos usuarios, genere una medida de similitud que permita de manera adecuada clusterizarlos, en función de sus historiales semánticos de recorridos.

Se incorpora como factor determinante para una buena clasificación de usuarios, considerar sus recorridos semánticos en vez de físicos. Por ejemplo: restaurant-> cine->café. Se mide entonces la similitud entre dos usuarios utilizando un algoritmo propuesto en el informe mencionado, de coincidencias máximas de viaje.

Solución Propuesta

La solución que se propone es tomar como base un informe realizado por la universidad de Hong Kong, para luego desarrollar el algoritmo que permita comparar usuarios, y así usar este valor como índice de comparación en algún algoritmo conocido de clustering.

Se tomaran datos de movimientos obtenidos en encuestas realizadas en la ciudad de Tandil, en diferentes familias, situadas en distintas zonas. Se desean incluir algunas variaciones a la solución propuesta por la Universidad de Hong Kong.

Para llevar a cabo este trabajo, se detallarán brevemente los aspectos principales que se deben tratar en esta tesis:

-Generar historial de movimiento: Para cada usuario que haya sido encuestado, generar su historial semántico de movimientos. Esto comprende tener una base de datos de lugares semánticos identificados en la región conocidos como POI (puntos de interés).

-Calcular la similitud entre cada par de usuarios, en base a su historial: Implementar el algoritmo “Maximal travel matches” que en base a dos usuarios y sus respectivos historiales semánticos, obtenga una medida de similitud entre los mismos, que permite ser incorporada a algún algoritmo de clustering.

- Clusterizar: Teniendo datos para utilizar como prueba, e implementada la funcionalidad requerida, aplicar clustering sobre los mismos.

Plan de trabajo

El plan de trabajo se detalla a continuación:

1. Análisis y relevamiento bibliográfico sobre minería de datos, clustering y funciones de similitud.
2. Comprensión del trabajo a partir del cual se pretende elaborar el algoritmo de comparación de usuarios.
3. Definición de regiones semánticas: El trabajo original comienza con recorridos conseguidos a través de GPSs, y luego se clasifican tomando POI (puntos de interés). En nuestro estudio, ya que los resultados serán obtenidos directamente como POI, debemos definir que POI serán tomados en cuenta, y en cuantos niveles de especificidad estarán (jerarquía semántica).
4. Almacenamiento y refinamiento de los resultados de las encuestas.

5. Implementación del algoritmo “Maximal travel matches”, que permite generar un valor de similitud, dados dos usuarios y sus historiales.
6. Clusterización de los usuarios, utilizando un algoritmo conocido de clustering, que use como función de similitud los resultados obtenidos luego de aplicar el algoritmo implementado.
 7. Análisis e interpretación de los resultados obtenidos, luego de clusterizar.
8. Escritura del informe final de tesis y conclusiones.

Bibliografía

[1] *Inferring Social Ties between Users with Human Location History*- Xiangye Xiao, Yu Zheng, Qiong Luo, Xing Xie. Microsoft Research Asia.

[2] *Techniques of Cluster Algorithms in Data Mining*- Johannes Grabmeier, Andreas Rudolph.

[3] *Friendship and Mobility: User Movement In Location-Based Social Networks*- Eunjoon Cho, Seth A. Myers, Jure Leskovec.

[4] *Social and Community Intelligence*- Daqing Zhang, Bin Guo, Zhiwen Yu.

[5] *Searching Trajectories by Locations: An Efficiency Study*- Zaiben Chen, Heng Tao Shen, Xiaofang Zhou, Yu Zheng, and Xing Xie.

[6] *Tutorial on Location-Based Social Networks*- Yu Zheng. Microsoft Research Asia.

[7] *Mining User Similarity Based on Location History*- Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, Wei-Ying Ma.

[8] *Finding Similar Users Using Category-Based Location History*- Xiangye Xiao, Yu Zheng, Qiong Luo, Xing Xie.

[9] *Location-based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data*- Jie Bao, Yu Zheng, Mohamed F. Mokbel.

[10] Recommending Friends and Locations Based on Individual Location History- Yu Zheng, Lizhu Zhang, Zhengxin Ma, Xing Xie, and Wei-Ying Ma.

[11] Learning travel recommendations from user-generated GPS traces- Yu Zheng, Xing Xie.

[12] Mining Interesting Locations and Travel Sequences From GPS Trajectories- Yu Zheng, Xing Xie, and Wei-Ying Ma.