IIMB Management Review (2016) 28, 25–30

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/iimb

IMR-DCAL SPECIAL PAPER

A study and analysis of recommendation systems for location-based social network (LBSN) with big data

Murale Narayanan [a,*], **Aswani Kumar Cherukuri** [b]

[a] *Information Technology, EMC Corporation, India Center of Excellence, Bangalore, India*
[b] *School of Information Technology & Engineering, VIT University, Vellore, Tamil Nadu, India*

Received 28 March 2014; revised 15 June 2015; accepted 22 January 2016; available online 13 February 2016

KEYWORDS
Big data;
Data mining;
Recommendation system;
Social network;
LBSN

Abstract Recommender systems play an important role in our day-to-day life. A recommender system automatically suggests an item to a user that he/she might be interested in. Small-scale datasets are used to provide recommendations based on location, but in real time, the volume of data is large. We have selected Foursquare dataset to study the need for big data in recommendation systems for location-based social network (LBSN). A few quality parameters like parallel processing and multimodal interface have been selected to study the need for big data in recommender systems. This paper provides a study and analysis of quality parameters of recommendation systems for LBSN with big data.

Introduction

Recommender systems or recommendation systems (RS) collect information based on the preferences of users (for example—songs, movies, jokes, books, travel destination and e-learning material). Recommender systems work based on users' information from different sources and provide recommendation of items. This information can be explicit (user rating) and implicit (monitoring user's behaviour), with millions of users using social networking services like Facebook, Twitter, and so forth. The rich knowledge that has accumulated in these social networking sites enables a variety of recommendation systems for its users.

A social network is an abstract structure comprised of individuals connected by one or more types of relations, such as friendship, shared knowledge, and common interests as stated by Zheng, Zhang, Xie, and Ma (2009). Location data add strength to the connection of the social networks. A location can be represented in relative, absolute, and symbolic form. Location is usually represented in three kinds of geographical representations—a point location, a region, and a trajectory.

In recent times, localisation techniques have enhanced social networking services, allowing the users to share their location-related content, and locations such as geo-tagged

[*] Corresponding author. Tel. +91 9972565848; fax: 41659511.
E-mail address: muralen@gmail.com (M. Narayanan).
Peer-review under responsibility of Indian Institute of Management Bangalore.

http://dx.doi.org/10.1016/j.iimb.2016.01.001

photos and notes. This is known as location-based social networks (LBSNs) (Zheng et al., 2009). An LBSN adds a location to an existing social network, and also tells the people in their social network that they can share their location-related information. Based on the location-related information, a new abstract structure is derived and connects connected individuals based on their location-related content, such as photos, texts and videos. Instant location and the history of a person are given as a timestamp during a certain period.

The advances in wireless communication technologies and location acquisition enables people to add a location dimension to traditional social networks and promotes a bunch of LBSN services, such as Foursquare, GeoLife and Loopt, where users can easily share their experiences in the physical world through mobile devices. The location dimension bridges the gap between the physical world and the digital online social networking services, giving rise to new opportunities and challenges in traditional recommender systems in the following aspects—complex objects and relations, and rich knowledge.

Location is one of the important components of user context and implies extensive knowledge about a user's interests and behaviour, thereby providing us with opportunities to better understand users in an abstract structure not only according to user behaviour, but the mobility of the user and his/her activities in the physical world. In recent times location-based services, such as tour guide and location-based social network, have accumulated a lot of location data. Today, the positioning function in mobile devices, such as GPS-phones, lets people know their locations easily. This location data provide various location-based services on the web and has shown itself to be attractive to the users. In real time, data are huge in volume, but data warehouses use small-scale datasets of users for recommendation.

When it comes to real-time scenario, these techniques may fail because millions of users will use social networks at the same time. The major challenges to be addressed in LBSN recommendation are 1) location-context awareness; 2) heterogeneous domain and 3) rate of growth.

Different types of data sources are used in recommendation systems for LBSNs, including 1) user profiles, 2) user online histories and 3) user location histories. This involves huge volumes of data in real-time scenario. Most recommendation systems in LBSNs currently use only one type of data source to make recommendations. Moreover, many of the data sources are related and may mutually reinforce each other. By considering more diversified data sources, more effective recommendations can be provided. For instance, the user online interactions, social structures and location histories are all very relevant to friend recommendation. If two users have more online interactions, are close in the social structure, and have overlapped location histories, these users are likely to be compatible. A friend recommender system that can consider all these factors will make higher quality friend recommendations.

We carried out an analysis, based on the characteristics of a recommender system, to give a comparison between big data and data warehouse with a dataset collected from Foursquare users, using a qualitative approach. This paper is organised as follows: Section 2 deals with literature review; Section 3 explains the challenges of the domain; Section 4 provides the objective of the paper; Section 5 details the dataset discussed in this paper; Section 6 gives characteristics of a location-based recommendation system; Section 7 explains the qualities of the location-based recommendation system, and Section 8 provides the conclusion.

Related work

Social media recommendations

Social media recommendation aims to provide users with suggestions of photos, videos, or other web content they might like. Using location information in LBSNs can improve both the effectiveness and efficiency of traditional social media recommendations. Several works in spatial keyword search for web content show the effectiveness of this pairing (Bouidghaghen, Tamine, & Boughanem, 2011; Cao, Cong, & Jensen, 2010b, 2011; Chen, Geyer, Dugan, Muller, & Guy, 2009; Zhang, Chee, Mondal, Tung, & Kitsuregawa, 2009). Location-aware image ranking algorithms have been proposed to increase the relevance of search results (Arase, Xie, Duan, Hara, & Nishio, 2009; Kawakubo & Yanai, 2011; Silva & Martins, 2011), which in turn improves the quality of the image tags, using a recommender system to automatically infer and suggest candidate location tags (Daly & Geyer, 2011).

The efficiency of recommendation systems can be significantly improved by using location data to prune out irrelevant information (Scellato, Mascolo, Musolesi, & CrowCroft, 2011). This improves the efficiency of content delivery networks using a novel caching mechanism based on geographic location. A real-time recommendation system, as suggested in Sandholm and Dung (2011), has been built for online web content using a collaborative filtering method to make more diverse and personalised recommendations within a geographical area. Levandoski, Sarwat, Eldawy, and Mokbel (2012) have proposed a novel location-aware recommendation system (LARS) framework to exploit users' ratings of locations using a technique that uses the distance of querying users to influence recommendations.

Categorisation by methodology

Although traditional recommendation systems have been successful by using community opinions, such as inventories in Amazon (Linden, Smith, & York, 2003) and news from Google (Das, Datar, Garg, & Rajaram, 2007) incorporating location information requires novel approaches. In this section, we categorise the major methodologies used by recommendation systems in location-based social networks as being based on — 1) content-based recommendation and 2) link analysis.

Content-based recommendations

Content-based recommendation systems, such as context aware and location based using Bayesian model (Park, Hong, & Cho, 2007; Ramaswamy et al., 2009), match user preferences discovered from users' profiles with features extracted from locations, such as tags and categories, to make recommendations. These systems require accurate and structured information for both the user profiles and the location features to make high-quality recommendations. The major advantage of the content-based approach in such a system is that it is robust against the cold start problem for both new users and locations. As long as the newly added users or locations have the appropriate descriptive content, they can be handled effectively. However, content-based recommendation systems have many drawbacks with regard to LBSNs: 1) content-based recommendation systems do not consider the aggregated community opinions (inferred from users), which may result in low-quality recommendations; and 2) content-based recommendation systems require that the structured information for both users and locations be created and maintained, which can be costly, especially in LBSNs in which the majority of the content (such as user profiles and location tags) is generated by the users.

Link analysis-based recommendations

Link analysis algorithms, such as PageRank (Page, Brin, Motwani, & Winograd, 1999), and Hypertext Induced Topic Search (HITS) (Chakrabarti et al., 1998; Kleinberg, 1999), are widely used to rank web pages. These algorithms extract high-quality nodes from a complex network by analysing the structure. In LBSNs, there are interconnected networks of different types, such as user–user, user–location and location–location networks. Zheng et al. (2009) extend the HITS algorithm for discovering experienced users and interesting locations in an LBSN. In their system, each location is assigned a popularity score, and each user is assigned a hub score, which indicates their travel expertise. Based on a mutually reinforcing relationship, a ranking of expert users and interesting locations is computed. Similarly, Raymond, Sugiura, and Tsubochi (2011) extend a random walk-based link analysis algorithm to provide location recommendation.

Challenges

Most recommendation systems in LBSNs currently use only one type of data source to make recommendations. As previously mentioned, there are different types of data in LBSNs. Many of the data sources are related and may mutually reinforce each other. By considering more diversified data sources, more effective recommendations can be provided. The recommendation methodologies used in the existing recommendation systems have their own drawbacks. For example, in collaborative filtering (CF)-based recommendation systems, data sparsity and cold starts are challenging problems. Link analysis-based recommendation systems avoid these problems, but only to provide generic recommendations that ignore users' personal preferences. By integrating CF and link analysis-based techniques, a hybrid recommendation system could overcome the weaknesses of both. For example, cold start problem arises due to insufficient amount of data for making reliable recommendations, while starting a recommendation system (Kleinberg, 1999). This problem can be overcome when we use big data that is a collection of large and complex datasets. The objective of this work is to address the previously discussed issues using big data and to report the impact of big data; where big data can store large volume of data and it can be accessed to get information about both user and location.

Objective of the paper

There are several approaches for a recommender system. Recommender systems use data warehouse bases to collect information about users. In real time, user information data are huge, so we are going to analyse a recommender system with big data. Data in a data warehouse are unmodified and consistent. Data formats are standardised through data warehouse to generate reports, and dashboards to forecast trends and predict the future. In data warehouse, data are integrated and provide interactive tools to users. Datasets for recommender systems play an important in LBSN to provide recommendations to users. To provide recommendation, data have to be available any time and need to be distributed. Our objective is to analyse whether data warehouse or big data is better for recommender systems. This can be done in two ways: qualitative and quantitative. Our analysis was qualitative, based on few parameters such as parallel processing and multimodal interface.

Characteristics of recommendation systems

Recommender systems should make use of the user's information, the user's friendships, location details, the user's interests, his/her social circles and the like. To extract this information, recommender systems would need to process structured and unstructured data laterally, which helps to discover unknown relationships from the data and extract information from multimodal interfaces. The present recommender systems use data warehouses for operation, but in real time, recommender systems have to recommend dynamically based on user information, do parallel processing, and handle structured and unstructured data as previously mentioned.

Many data warehouses have components that sense probable customer churn and use a recommendation system to convince the consumer to stay. Using predictive analytics and user profiles, personalised offers are sent proactively to potential users to earn their constant loyalty. In many industries, this application is mandatory. Big data helps in pulling click streams from various websites to discover consumer preferences/interests. This is also helpful to get churn results, and preference/interest data of the user.

The data warehouse and big data are not highly differentiated in some categories. And so, either tool could be the right solution. Choosing the best tool depends on the requirements. Further, big data and the data warehouse can work together in an information supply chain.

The analysis of a recommender system can be done using a qualitative or quantitative approach. This study analyses RS qualitatively. We took up few important parameters to analyse recommender systems using big data and data warehouse. Some parameters are explained below.

Structured and unstructured data

The data extracted from users and the source for recommender systems consist of structured and unstructured data. Data warehouse does not handle unstructured data, which is an important source for a recommender system. This is accomplished by using big data.

Distributed or centralised data

Recommender systems need distributed data to suggest offers to the users based on their interests, but in data warehouse, all data are centralised. Massive data sets are used to study the location of the user, but these data do not fit in data warehouse. Big data has been a boon to these approaches.

Data mining

The data mining algorithm should run in parallel to provide quick results. Big data runs predictive analytics in parallel against huge quantities of data. The data warehouse uses integrated data where big data often has raw data in quantity. Therefore, one way to choose between big data and the data warehouse for data mining is based on the data itself.

Parallel processing

Recommender systems depend on running complex jobs to process huge amounts of data. When the request has to run in parallel to achieve scalability and the programme is highly complex, big data has many advantages. Thus, any system that has to be run parallelly favours big data.

Multimodal interface

Multimodal data access (such as audio, video and clicks) is used for interaction data, users' affective responses, and contextual information; and exploits this information to provide meaningful recommendations. These unstructured data cannot be stored in a data warehouse. Thus, multimodal interface favours big data.

Dataset

We chose Foursquare[1] dataset of David Floyer (2013), the most popular LBSN, to study the user's check-in behaviour and social-historical ties on LBSNs. In the Foursquare dataset, we get a user's check-in history with timestamps and his/her friendship information. To collect user check-ins, Foursquare does not have any public application programme interface (API). So, we were not able to get the check-in history directly. However, users in Foursquare can choose to list on their respective Twitter accounts and publish their check-in messages as tweets on Twitter, and these can be accessed through Twitter's public API. This contains a unique URL that points to a Foursquare web page, including the geographical information of the user's check-in location. We obtained check-ins with timestamps ranging from August 2010 to November 2011. To keep the friendships identical to the Foursquare data, the Foursquare user's social circle was directly used from Foursquare. In our experiment, we considered users who had at least 10 check-ins. We obtained 43,108 unique geographical locations as the location vocabulary. In this dataset, the user's location is stored in terms of latitude and longitude. Sample data of the dataset are provided in Table 1 and Table 2.

Table 1 Sample data of Foursquare check-ins.

User ID	Latitude	Longitude	Time	Location ID
0	37.80617	−122.45	14-04-2010 04:32	0
0	37.80635	−122.448	14-04-2010 04:55	0
0	37.80396	−122.449	14-04-2010 16:06	0
0	37.77354	−122.409	15-04-2010 00:59	1
0	37.76174	−122.431	15-04-2010 02:59	2
0	37.7612	−122.431	15-04-2010 06:28	2
0	37.76163	−122.431	15-04-2010 15:07	2
0	37.7611	−122.433	15-04-2010 15:11	2
0	37.80818	−122.432	15-04-2010 17:07	3
0	37.80502	−122.433	15-04-2010 17:09	4
0	37.80803	−122.431	15-04-2010 17:14	3
0	37.80792	−122.431	15-04-2010 17:54	3
0	37.80817	−122.432	15-04-2010 18:36	3
0	37.80829	−122.432	15-04-2010 18:37	5
0	37.80886	−122.416	15-04-2010 18:39	6
0	37.80829	−122.432	15-04-2010 18:40	5
0	37.80829	−122.432	15-04-2010 18:42	5
0	37.8083	−122.431	15-04-2010 18:42	5
0	37.80503	−122.434	15-04-2010 18:51	4
0	37.80503	−122.434	15-04-2010 20:50	4
0	37.80753	−122.431	15-04-2010 20:54	3
0	37.80753	−122.431	15-04-2010 20:54	3
0	37.80886	−122.416	15-04-2010 21:51	6
0	37.80886	−122.416	15-04-2010 22:00	6
0	37.78139	−122.4	16-04-2010 00:01	1
0	37.79041	−122.39	16-04-2010 04:29	7
0	37.77837	−122.406	16-04-2010 05:16	1
0	37.77805	−122.406	16-04-2010 05:18	1
0	37.77858	−122.406	16-04-2010 06:45	1
0	37.78452	−122.404	16-04-2010 17:34	1
0	37.78237	−122.401	16-04-2010 18:21	1
0	37.8058	−122.267	17-04-2010 02:59	8
0	37.80772	−122.27	17-04-2010 03:25	9
0	37.80811	−122.27	17-04-2010 03:34	9
0	37.8081	−122.27	17-04-2010 03:46	9
0	37.80799	−122.27	17-04-2010 03:54	9
0	37.76519	−122.397	17-04-2010 04:30	10
0	37.76492	−122.396	17-04-2010 04:31	10
0	37.76505	−122.396	17-04-2010 04:34	10
0	37.76063	−122.432	17-04-2010 15:06	2
0	37.75872	−122.417	17-04-2010 16:15	11

The dataset information is as follows: no. of users: 18107; no. of check-ins: 2073740 and no. of links: 115574.

This dataset contains information about: check-in information of Foursquare users; friendship network of Foursquare users; and home location of Foursquare users.

Qualities of recommender systems

A qualitative study for recommender systems was carried out based on the parameters mentioned previously. We tried to

[1] Foursquare is a local search and discovery service mobile app that provides users search results, by considering the places visited by the user and things that the users like, as data in the app. This app provides recommendations to the users about places to visit.

Table 2 Sample data of Foursquare check-ins.

User1	User2
16401	6455
16401	685
16401	13767
16401	6720
16401	10755
16401	605
6455	16401
685	16401
685	13767
685	17836
685	12039
685	13792
685	2704
685	4740

Table 3 Qualities of recommender systems.

Parameters	Data warehouse	Big data
Structured or unstructured data	NA	++
Distributed or centralised data	NA	++
Parallel processing	NA	++
Multimodal interface	NA	++
Unrestricted or ungoverned explorations	NA	++
Central processing unit (CPU) intense analysis	++	++
Discover unknown relationship between data	++	++
Interactive reports and online analytical processing (OLAP)	++	NA
Security	++	NA

++ Available; NA, not available.

study certain use cases that were distinctive to big data or the data warehouse; there was also overlap where either technology could be efficient. We have taken a few quality parameters to study and analyse the big data in recommender systems. The comparison of the quality parameters with big data and data warehouse is provided in Table 3.

The qualitative comparative study favours the big data approach because:

- The recommender system data is distributed through many systems.
- The data schemes are simple and flat, which helps to establish the user experience.
- The availability and quality of data were indefinite at the beginning and needed much iteration before the appropriate data could be selected and transformed.
- Large amounts of data need to be extracted. It was not possible to centralise the data for recommender systems.
- Even financially the big data approach is more efficient than data warehousing, i.e. break even months are fewer for big data.

Conclusion

Till date, recommender systems make recommendations only by getting data from data warehouses. Due to the dynamic nature of the recommender system, the data have to be distributed. The recommender system has to work parallelly to provide recommendation to the user and to support different interfaces. Big data excels in handling unstructured, raw and complex data with huge programming flexibility. This study analyses the use of big data in recommendation systems qualitatively. In future research, we will attempt to analyse the recommendation system in social networks quantitatively.

References

Arase, Y., Xie, X., Duan, M., Hara, T., & Nishio, S. (2009). A game based approach to assign geographical relevance to web images. In *Proceedings of the 18th International Conference on World Wide Web* (pp. 811–820). ACM.

Bouidghaghen, O., Tamine, L., & Boughanem, M. (2011). Personalizing mobile web search for location sensitive queries. In *Mobile Data Management (MDM)* (Vol. 1. IEEE, pp. 110–1182011). 12th IEEE International Conference.

Cao, X., Cong, G., & Jensen, C. (2010b). Retrieving top-k prestige-based relevant spatial web objects. In *Proceedings of the VLDB Endowment* (p. 3, 1–2, 373–384).

Cao, X., Cong, G., & Jensen, C. (2011). Collective spatial keyword querying. In *Proceedings of the 2011 International Conference on Management of Data* (pp. 373–384). ACM.

Chakrabarti, S., Dom, B., Raghavan, P., Rajagopalan, S., Gibson, D., & Kleinberg, J., (1998). Automatic resource compilation by analyzing hyperlink structure and associated text. Computer Networks and ISDN Systems. (pp. 30, 1–7, 65–74).

Chen, J., Geyer, W., Dugan, C., Muller, M., & Guy, I. (2009). Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 201–210). ACM.

Daly, E., & Geyer, W. (2011). Effective event discovery: using location and social information for scoping event recommendations. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (pp. 277–280). ACM.

Das, A., Datar, M., Garg, A., & Rajaram, S. (2007). Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 271–280). ACM.

Floyer, D., (2013). Financial comparison of big data MPP solution and data warehouse appliance. Wikibon Article.

Kawakubo, H., & Yanai, K. (2011). Geovisualrank: a ranking method of geo-tagged images considering visual similarity and geo-location proximity. In *Proceedings of the 20th International Conference Companion on World Wide Web* (pp. 69–70). ACM.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46*(5), 604–632.

Levandoski, J., Sarwat, M., Eldawy, A., & Mokbel, M. (2012). Lars: a location-aware recommender system. In *IEEE International Conference on Data Engineering*.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *Internet Computing*, IEEE, *7*(1), 76–80.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999), The page rank citation ranking: bringing order to the web. Stanford Technical Report.

Park, M., Hong, J., & Cho, S., (2007). Location-based recommendation system using Bayesian user's preference model in mobile devices. Ubiquitous Intelligence and Computing, (pp. 1130–1139).

Ramaswamy, L., Deepak, P., Polavarapu, R., Gunasekara, K., Garg, D., Visweswariah, K., et al. (2009). Caesar: a context-aware, social recommender system for low-end mobile devices. In *Mobile data management: systems, services and middleware, 2009. MDM'09. Tenth International Conference on IEEE* (pp. 338–347).

Raymond, R., Sugiura, T., & Tsubochi, K. (2011). Location recommendation based on location history and spatio temporal correlations for an on-demand bus system. In *ACM SIGSPATIAL*. ACM.

Sandholm, T., & Dung, H. (2011). Real-time, location-aware collaborative filtering of web content. In *Proceedings of the 2011 Workshop on Context-Awareness in Retrieval and Recommendation* (pp. 14–18). ACM.

Scellato, S., Mascolo, C., Musolesi, M., & CrowCroft, J. (2011). Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 457–466). ACM.

Silva, A., & Martins, B. (2011). Tag recommendation for georeferenced photos. In *Proceedings of the 3nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*. ACM.

Zhang, D., Chee, Y., Mondal, A., Tung, A., & Kitsuregawa, M. (2009). Keyword search in spatial databases: towards searching by document. In *IEEE International Conference on Data Engineering* (pp. 688–699). IEEE.

Zheng, Y., Zhang, L., Xie, X., & Ma, W. (2009). Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th International Conference on World Wide Web* (pp. 791–800). ACM.