ARPN Journal of Engineering and Applied Sciences

SERVICE RECOMMENDATION SYSTEM IN SOCIAL NETWORKS

Sandra Elizabeth Salim and R. Jebakumar
Department of Computer Science and Engineering, SRM University, Chennai, India
E-Mail: sandraelizabeth.05@hotmail.com

ABSTRACT

Social networks have become an inevitable part of today's life. The content from social media can be used for a number of purposes; one of the main being recommendation systems (RS). Traditional recommendation systems ignored the concept of social media and its influence on people. There have been a lot of RS in industry since the last decade. In this paper, we have proposed a new system called Service Recommendation in Social Networks (SRSN) based on a keyword approach which overcomes the drawbacks of current generation of RS. It utilizes the concept of user based collaborative filtering algorithm (UCF) in generating recommendations. SRSN is designed to work in big data environments as it is implemented in Hadoop, a map reduce paradigm.

Keywords: recommendation system, big data, social network, collaborative filtering.

1. INTRODUCTION

Social networks have provided a new platform for people to interact and share information. Social networks are now the best medium to create public awareness and spread useful information to the people. Nowadays, numerous sources deliver information in different forms and this data is mostly produced from social network site users through their reviews and comments [1]. People find this type of content more useful than those produced by professional writers. Before the era of internet, it was difficult for people to take decisions regarding what service to go for by evaluating the pros and cons of these services. The only option was to get opinions and suggestions from their friends and relatives. With the advent of internet and especially social media, this has become an easy task. People can easily get any amount of information about a particular service through internet and social media [2].

In some cases, users find it difficult to exactly judge what service to go for or may be what to buy; they may not get the exact result what they are looking for. Some people resort to get help from multiple blogs, news articles or events which are related to the item they are searching for. But still in this case, users need to visit various sources and go through them to differentiate between the content required by them and content irrelevant to them. In such cases, the best way is to provide recommendations to the users regarding their exact requirement.

Users feel more comfortable and satisfied to get recommendations from their friends, relatives and other known people than getting from strangers and people unfamiliar to them. This is accomplished through social networks [3]. There are a number of social networking sites such as Facebook, Twitter, Quora, LinkedIn, etc. Through these networks, we get a system where information is integrated from multiple sources to provide content and data to users sharing common interests and ideas. Users can give feedbacks and suggestions about various services which are useful for all other members in that particular social media or communities within the network.

However, the current generation of RS have many drawbacks and requires further improvements to make recommendation methods more effective. These improvements include better methods for representing user characteristics and the information about the items to be recommended, incorporation of various user given contextual information into the recommendation process, utilization of multicriteria ratings, development of less intrusive and more efficient recommendation methods for the big data environment.

In this paper, we propose a new technique which overcomes the shortcomings of existing recommendation systems in social networks. In Section 2 and 3, we provide a small introduction about social media and big data along with the current recommendation systems. In Section 4, we propose a new system, SRSN which uses a keyword based approach for generating recommendations.

2. SOCIAL MEDIA

Social networks have provided an advanced method for people to communicate with each other, that it has become a part and parcel of today's modern era [4]. Nowadays, there are hundreds of sources delivering content in various forms and more content is generated from users through social networks and the reviews and comments in these networks than through professional writers. Users are often faced with information overflow. In the days before Internet, people found it hard to determine what book to read, what movie to watch, or which place to visit, etc. They were often guided with the recommendations and feedbacks of their friends.

Social media enhances social cooperation and communication among people where they exchange information and share ideas [5]. Social media has become an inevitable aspect of today's life. We hardly find anyone who is not aware or who is not a part of any social network. Social media plays an important role in bridging the communication gap between organizations and individuals.

Social media utilizes mobile and web-based features to form advanced interactive environments through which people communicate, share and create their

VOL. 10, NO. 10, JUNE 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ideas. This technology has introduced considerable changes in the way in which organizations and individuals communicate. Earlier, people used to receive information and data from electronic and print media. In this regard, social media is distinct in the aspect that it can be accessible anywhere by anybody and at the same time inexpensive. It provides a platform where anybody is free to share whatever he/she feels like and forms an excellent medium for communication [6]. The main feature of social media is its ability to reach out even to the smallest sections in society. News through social media spreads even faster than a forest fire. Figure-1 shows a modelling diagram of how social media is related to RS.

It has been estimated by IBM that that 2.5 quintillion bytes of data are created each day. It has been noted that social media alone creates million bytes of data in a less amount of time than compared to the data production which occurred in the world previously [7]. The data and information created by these social networks is not only enormous and huge - but also unformulated. The crucial job of managing, processing and monitoring this data undoubtedly exceeds human ability. It's even beyond the scale of most common software. It's easy to understand how Big Data fits into the picture. The Big Data industry deals with large sets of data that range from terabytes to many hundreds of petabytes. In order to utilize it and turn it into useful information, there is a need for us to start turning to Big Data for translation [8].

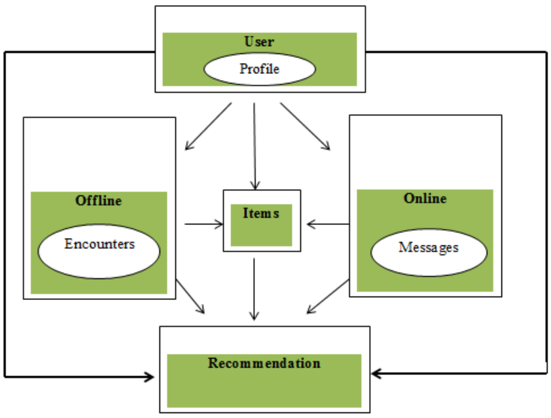


Figure-1. Modelling social media recommendation.

2.1 Big data

Big data includes large and enormous amount of data which are complex and are hard to process using normal and traditional data processing applications. There are a lot of challenges in processing big data applications such as analysis, search, capture, sharing, curation, storage, privacy violations, etc. Big data is now rapidly developing in all fields of science and engineering domains [9]. The main concepts of big data include its large-volume, autonomous and heterogeneous data sources with decentralized and distributes control. These features form a big challenge for harnessing data. An important area where big data comes into play is social media. Quintillion bytes of data are generated every day in social networks alone [10].

For the most important big data applications such as Google, Facebook, Flip kart, etc. there are a lot of server farms which are deployed all over the world in order to ensure quick and nonstop responses for user queries [11]. It is difficult for big data to work with normal relational database management systems and visualisation packages.

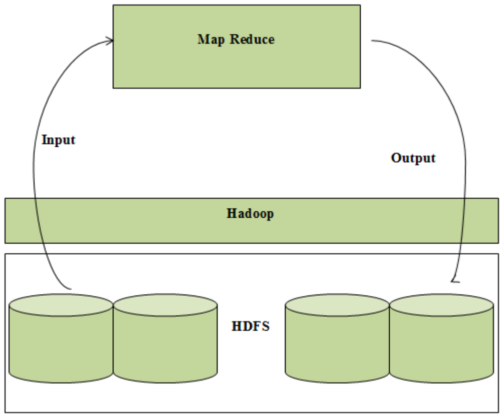


Figure-2. Map reduce paradigm.

In 2004, Google published a paper on a process called MapReduce. This framework provides a parallel processing model along with its associated implementation with regard to big data applications. Figure-2 depicts the most important phases in map reduce paradigm using Hadoop. In map reduce paradigm, queries are split and given across node which are processed in parallel [12]. The results are then gathered and delivered. Map reduce thus has two phases: the map phase and reduce phase. Map reduce framework became very successful and so others wanted to replicate and develop the algorithm. An implementation of map reduce was undertaken by Apache open source project and was named as Apache Hadoop.

Big data gained popularity all of a sudden and need for big data mining has been arising in almost all domains of science and engineering. With this technology, we will be able to provide most accurate social sensing and most relevant feedback for better understanding of society [13]. We can also increase the participation and contribution of public audiences for social and economic events.

3. RECOMMENDATION SYSTEMS

Recommendation systems are an important type of Information Filtering method. Recommendation techniques have been studied widely under data mining, machine learning and information retrieval. We use the term - *item* to represent the thing which system recommends for the user. In order to make daily decisions regarding products and services, we often depend on recommendations provided by other people [14]. It is from this concept that the idea of RF was initiated. Traditional recommendation systems never consider the social relationships among users. Integrating RS with social content can increase the performance and efficiency of these systems.

Recommendation systems were developed with the intension of helping people in the process of decision making process where they had less personal experience and for people who found it difficult to evaluate the overwhelming number of items that a website offers. Individuals rely on opinions and reviews of other people in making daily decisions [15]. The development of Internet and associated technologies posed a threat for traditional RS. Traditional recommendation systems never took into account the aspect of social recommendations. But in real life we are actually resorting to social recommendations where we ask for recommendations from our near and dear ones [16]. Social recommendation is a daily occurrence.

The most important function of a RS can be regarded as correctly identifying the core item which a user wants. The system must be accurately able to predict what the user wants and must be able to satisfy the user. RS are basically divided into the following types [17]:

- **Content-based:** The similarity of items is measured based on the characteristics of the items being compared.
- **Collaborative filtering:** The similarity of items is measured based on the rating history of users.
- **Hybrid recommendation systems:** The features of content based and collaborative filtering methods were combined.

4. SERVICE RECOMMENDATION SYSTEM USING KEYWORD APPROACH

Here, we propose a Keyword Based Recommendation System called service recommendation system in social networks (SRSN). SRSN uses a keyword-aware method for generating recommendations. It is based on user-based Collaborative Filtering algorithm (UCF).It uses reviews from previous users to retrieve user's choices and provides multi criteria of candidate services. To improve its efficiency and scalability in "Big Data" environment, we implement it in Hadoop, which is a MapReduce framework. The main steps of SRSN are shown in Figure-3 and described in detail as follows:

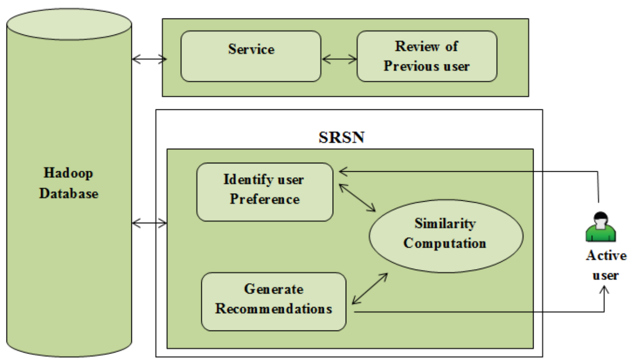


Figure-3. SRSN Architecture.

4.1 Identify user preferences

In this step, we identify the preferences of the active user. The proposed system has two sets of users: the previous user set and active user set.

Previous users will have two lists of keywords: possible keyword list (PKL) and a synonym list (SL). PKL is the one which stores the keywords extracted from the reviews of previous users. The previous users can give their feedbacks and reviews about a particular item/product in the respective social media site which they prefer. The keywords in these reviews are extracted and stored in PKL. The keywords from the reviews of previous users can be extracted using a number of ways. Here, we use the Alchemy API to extract keywords. Alchemy API provides easy to use facilities for extracting keywords. These URL processing calls automatically fetch the desired webpage, normalize it and extract the keywords. The "possible keyword list" is represented as PKL ={pk1,pk2,pk3…pkn}.

There can be situations where the active user gives preference with keyword as same meaning as that of the ones in PKL but the word is different. In order to provide a correct match of words in such situations, the synonyms of the words in PKL are stored in a separate list called "synonym list" (SL).

The active user is the one who actually needs recommendation for a particular item/product. He can give his choices from a list of keywords called the "keyword service list" (KSL). This list contains list of services for which recommendation is required. It is represented as KSL = {sk1, sk2, sk3, …, sk3}. After selecting the keyword from the list, the active user must also select the importance degree of the keyword. The importance degree is depicted by numbers from 1 to 5, where 1 represents general and 5 represents very important. Table-1 summarizes the basic symbols used in this paper.

Table-1.

S. No.	Symbol	Definition
1.	PKL	Possible Keyword List
2.	KSL	Keyword Service List
3.	SL	Synonym List
4.	JNV	Jaccard, N-Gram, Vector Space
5.	IR	Individual Rating
6.	CS	Candidate services
7.	KP	Previous user keyword
8.	DS	Data Source

4.2 Compute similarity

Here, we identify the reviews of previous users which have similar preferences to an active user. For this, we have to compute the similarity between the keywords in the PKL of previous user and the keyword chosen by the active user. Before computing the similarity between keywords, the reviews of the previous users which are not

ARPN Journal of Engineering and Applied Sciences

related to the preferences of active user must be filtered out. This is done using the intersection concept in set theory. If the intersection between KSL of active user and PKL previous user is an empty set, then the "possible keyword set" of previous user will be filtered out.

The most common methods used for computing similarity between keywords are Jaccard's, N-Gram and Vector Space model [18]. Each of these methods has their own pros and cons. In order to get the best outcome, we use a hybrid method which combines all the three methods. Similarity between two keywords using Jaccard's coefficient [19] is done using the following equation:

$$J\ (A,B) = \frac{A \cap B}{A \cup B} \tag{1}$$

where

A = set of characters that doesn't occur in word in KSL
B = set of characters that doesn't occur in words in PKL_j

In N-Gram model the probability of character sequence that occurs as a word is calculated [20]. The size of N-Gram can be from 1 to (n). Similarity of words using this model is computed using equation (2). Here, the lengths of character and word sequences are different.

$$NG\ (2,X)= \{x_0x_1, x_1x_2, x_2x_3, \ldots x_{n-1}x_n\}$$
$$NG\ (3,X)= \{x_0x_1x_2, x_1x_2x_3, \ldots x_{n-2}x_{n-1}x_n\} \tag{2}$$

In Vector Space model [21], keywords are represented in forms of t-dimensional Space Vector of word weighting. The similarity is computed using equation (3).

$$VS\ (A,B)= \frac{VA}{Len(A)} * \frac{VBj}{Len(Bj)} \tag{3}$$

where

V_A= Vector of number of characters in word in KSL
V_{Bj}= Vector of number of characters in words in PKL_j
Len (A) = Length set of all characters in word in KSL
Len $(B)_j$= Length set of all characters in words in PKL_j

We use Jaccards N-Gram Vector Space Average (JNV) method which is created by combining the three methods;the computation is done using equation (4).

$$JNV(KSL,PKL_j) = \frac{J\ (KSL,PKLj) + NG(KSL,PKLj) + VS(KSL,PKLj)}{3} \tag{4}$$

Here, we use the reviews of previous users to suggest recommendations to active users, and hence collaborative filtering comes into play. Thus, the similarity measure is done using a combination of collaborative filtering and JNV method. The similarity between keywords k_i and KP_j is computed using formula 1. Algorithm SC (Similarity Computation) illustrates the similarity computation method. The similarity between keywords k_i and KP_j is computed using formula 4. Here k_i denoted keywords from candidate services (RCS) and KP_j denotes keywords in PKL of previous users.

Algorithm SC (SRSN, PKL, RCS)
1. for each $KP_j \in$ PKL
2. for each $k_i \in RCS$
3. if $k_i \in KSL$
4. JNV (k_i, KP_j) = $\frac{J\ (ki,KPj) + NG\ (ki,KPj) + VS(ki,KPj)}{3}$
5. end if
6. end for
7. end for
8. return JNV (k_i, KP_j)
9. End SC

4.3 Generate recommendation list

In this step, further filtering of previous user keywords will be conducted depending on the similarity of active user and previous users. If the similarity between KSL and PKL_j is less than a particular threshold value∂, PKL_j will be filtered out, otherwise PKL_j will be retained. After the set of similar users is identified, the individual rating of each candidate service is calculated. The individual rating (IR) of a service is calculated using a weighted average approach as shown in equations 5 and 6. Here, JNV $(\mathrm{KSL}, \mathrm{PKLj})$is the similarity of preference keyword set of active user KSL and preference keyword set of a previous user PKL_j; multiplier k serves as a normalizing factor; S denotes the set of the remaining preference keyword sets of previous users after filtering; r_j is the rating of the corresponding review of PKL_j, and s is defined as the average ratings of the candidate service. At the end, a personalised service recommendation list is provided to the active user. The services which have the highest ratings will be given as output

$$IR = s + k \sum_{KSL \in S} \mathrm{JNV}\ (\mathrm{KSL}, \mathrm{PKLj}) * (rj - s) \tag{5}$$

$$k=1/\sum_{KSL \in S} \mathrm{JNV}\ (\mathrm{KSL}, \mathrm{PKLj}) \tag{6}$$

The individual ratings of all services for an active user could be computed by following the steps described above. The services are ranked according to the rating which is presented to the active user. Finally, the services with top K ratings are provided to the user. Algorithm SRSN illustrates the complete functionality of the system.

Algorithm SRSN (DS, RCS, KSL,∂, K)
1. Total = 0, r = 0, S = NULL
2. PKL = NULL
3. for each RCS (j) do
4. KP = A (RCS)
5. PKL = PKL ∪ KP ∈ DS
6. if $PKL_j \cap$ KSL ≠ NULL, then
7. S = S ∪ PKL_j
8. end if
9. for each keyword $PKL_j \in$ S
10. sim = SC (KSL, PKL, RCS)
11. if (sim) <∂, then

ARPN Journal of Engineering and Applied Sciences


```
12.                         S = S - PKL_j
13.                         else
14.                         total = total + 1, r = r + r_j
15.              end if
16.      end for
17. s = r/total
18. IR = s + k ∑_{KSL∈S} sim * (rj − s)
19. sort S based on IR
20. return top K values in S
21. End SRSN
```

The input to this algorithm consists of a data source (DS) which can be any social network site or blog, reviews of candidate services (CS), possible keyword list (PKL) showing the current user's preference, a threshold value ∂, and the number K which is used to return the top K ratings. S is used to store the remaining keywords in PKL and *total* is used to record the number of remaining keywords in PKL. Step 4 is used to extract keywords from previous user reviews using Alchemy API into previous user keyword list (KP). PKL is initiated with words in KP in step-5. In steps 6 to 8, filtering process is performed. If intersection between PKL and KSL is not NULL, the keywords in PKL are entered into the set S. Similarity computation and further filtering is done in steps 9 to 16. Similarity between KSL and PKL_j is computed using JNV (KSL, PKL_j). If the result is less than a particular threshold value (∂), then PKL_j will be removed from S, else it is retained in S and the value of *total* and *r* are computed. In steps 17 and 18, the average rating (s) and individual rating (ir) is computed. Finally in steps 19 and 20, the set S is sorted based on the value of IR and the result is returned with top K ratings to the current user.

5. IMPLEMENTATION

SRSN is implemented in a Map Reduce framework on Hadoop platform. This is done so as to improve the scalability and efficiency of the system. Map Reduce framework is used to process large amount of data. Thus, the efficiency of SRSN will not go down even if the size of data is large. Implementation of SRSN in Hadoop mainly consists of three steps.

Step-1: The first step is to extract keywords (PKL) from reviews of previous users for each services and compute the average rating (s) for each service. The reviews of previous users are given as input to MapReduce1 ().

Step-2: The similarity between active and previous users is computed in this step. The output of MapReduce1 () along with KSL is given as input to MapReduce2 (). The similarity is computed using Algorithm SC (PKL, KSL, RCS).

Step-3: In this step, MapReduce3 () is used to calculate the individual rating of each service and generate a personalized recommendation list for the active user (equation 5 and 6). The output of MapReduce2 () is given as input to MapReduce3 (). Finally MapReduce3 () outputs the top K recommendations to active user.

6. EXPERIMENTAL EVALUATION

In this section, we perform various experiments to evaluate the accuracy and precision of SRSN. To evaluate these factors, we use 2 metrics: Matthews Correlation Coefficient (MCC) and F1 Score (weighted average of precision and recall). SRSN can be implemented in any social media, blogs, events, etc. Here, we use three data sets as shown in Table-2. These experiments are conducted in a Hadoop platform.

Table-2.

Data set 1 (DS 1)	Data extracted from Facebook comments and posts
Data set 2 (DS 2)	Data extracted from blog.laptopmag.com
Data set 3 (DS 3)	Data extracted from imdb

6.1 Experimental setup

MCC takes into account true and false positive and negative values. MCC always returns a value between -1 and +1. A value of +1 indicates a perfect prediction. The F1 score is the harmonic mean of precision and recall. In our context, precision is the proportion of recommendations that are good and recall is the proportion of good recommendations that appear in top recommendations. A value of 1 shows the best outcome of F1 score.

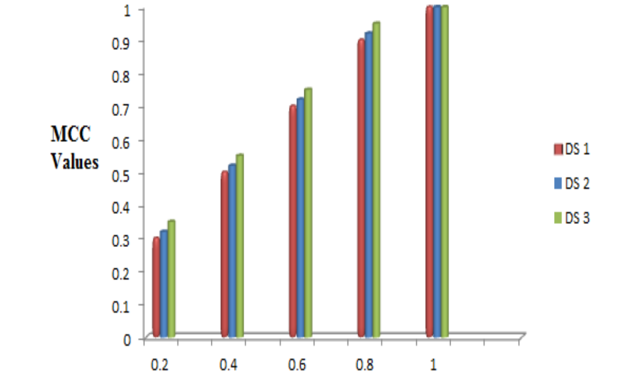
Figure-4 plots a graph between False Positive Rate and True Positive Rate of the three data sets whereas Figure-5 plots a graph between the values of precision and recall. As is evident from the figure, both tables 3 and 4 exhibit ideal values of MCC and F1 score. MCC values are computed from False Positive Rate and True Positive Rate. F1 score is interpreted from the values of precision and recall. Table-4 shows the MCC values and F1 scores of the three data sets DS 1, DS 2 and DS 3.

Table-3.

Data Set	MCC	F1 score
DS 1	1	0.98
DS 2	0.96	1
DS 3	1.02	1.1

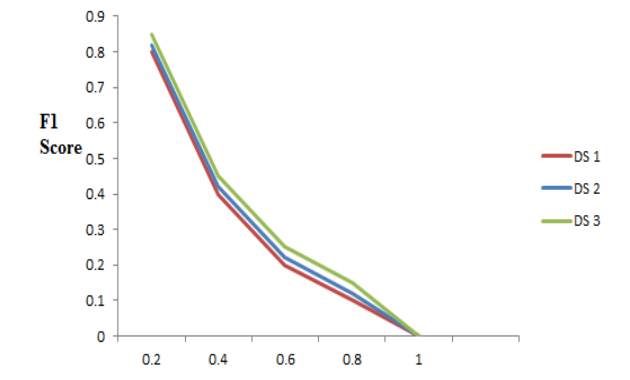
Table-4.

Data Set	MCC	F1 score
DS 1	1	0.98
DS 2	0.96	1
DS 3	1.02	1.1



DS 1	0.301	0.502	0.702	0.901	1.10
DS 2	0.322	0.510	0.722	0.932	1.01
DS 3	0.351	0.552	0.756	0.962	1.00

Figure-4. MCC values.



DS 1	0.812	0.401	0.212	0.101	0.01
DS 2	0.821	0.432	0.243	0.123	0.21
DS 3	0.857	0.456	0.253	0.153	0.03

Figure-5. Precision-recall values.

7. CONCLUSION AND FUTURE WORK

Over the last few decades, a lot of developments have been made in the field of recommendation systems. In spite of these advancements, the current generation of recommendation systems still requires a lot of advancements in order to improve its efficiency and accuracy. Existing recommendation systems hardly focused on social relationships among people. In reality, when we ask for suggestions from our friends and relatives, it is actually a form of social recommendation. Thus, in order to enhance recommendation systems and to deliver more personalized and accurate recommendations, incorporating social media information among users have become inevitable.

In this paper, we have proposed a recommendation system called Service Recommendation in Social Networks (SRSN) which uses social media information to provide the required results. Here, the user's preferences are indicated through keywords and recommendations are generated using a user-based Collaborative Filtering algorithm. A keyword list is provided to choose user's preferences. The active user can give their choice by selecting the options from the keyword-service, and reviews of existing users are extracted from their ratings. SRSN presents a personalized and accurate service recommendation list to users by incorporating multi criteria rating. In addition to that, SRSN is implemented in Hadoop in order to increase the efficiency and improve the scalability of the entire process in Big Data environment. In future, we will do further research on how to identify and extract the user's feedback from pictures and video. At present only textual representation is incorporated.

REFERENCES

[1] Fu XJ, Goh RSM, Tong JC, "Social Media for Supply Chain Risk Management", IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2013.

[2] M D Dawot, Ibrahim R, "A Review of Features and Functional Building Blocks of Social Media", 8th IEEE International Conference on Software Engineering, 2014.

[3] Hian Hu, Yonggan Wen, Huanbo Luan, "Towards Multiscreen Social TV with Geolocation Aware Social Sense", IEEE Transactions of Multimedia 2014, Volume 21, Issue 3.

[4] King I, Jiexing Li, Kam Tong Chan, "A Brief Survey of Computational Approaches in Social Computing", IEEE International Conference on Neural Networks, 2009.

[5] Jung Tea Kim, Jong Hoon Lee, Sangwook Kim, "Social Contents Sharing Model and System Based on User Location and Social Network", IEEE Third International Conference on Consumer Electronics, 2013.

[6] Hoon Ki Lee, Eui Hyun Paik, Juan Tee Kim, "Location Based Personalized Social Media Service for the Smart Phone: Distributed Social Contents Rating", IEEE International Conference on Internet Technology and Secured Transactions (ICITST), 2013.

[7] Ying Chiang Cho, "Violence and Aberration in the Age of Social Media: Transforming the Advanced Communication Technology into a Better Facilitative Tool", IEEE Transactions on Consumer Electronics 2014, Volume 3, Issue 4.

[8] Zhaohui Wu, OoiBengChin "From Big Data to Data Science: A Multi-disciplinary Perspective", Big Data

ARPN Journal of Engineering and Applied Sciences

Research 1 (2014) 1, 2214-5796/©2014 Published by Elsevier Inc.

[9] Laurence T. Yang, Jinjun Chen “Special Issue on Scalable Computing for Big Data”, Guest Editorial / Big Data Research 1 (2014) 2–3, 2214-5796/©2014 Published by Elsevier Inc.

[10] Demchenko Y, de Laat C, Membrey P, “Defining Architecture Components of the BigData Ecosystem”, IEEE International Conference on Collaboration Technologies and Systems (CTS), 2014.

[11] Lei Wang, Jianfeng Zhan, ChunjieLuo, “BigData Bench: A BigData Benchmark Suite from Internet Services”, IEEE International Symposium on High Performance Computer Architecture (HPCA) 2014.

[12] Chia-WeiLee, Kuang-YuHsieh, Sun-YuanHsieh, Hung-ChangHsiao, “A Dynamic Data Placement Strategy for Hadoop in Heterogeneous Environments”, Big Data Research 1 (2014) 14–22, 2214-5796/©2014 Elsevier Inc.

[13] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, “Data Mining with Big Data”, IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 1, January 2014.

[14] Yongjian Fan, Yanguang Shah, “Study of the Model of E-Commerce Personalized Recommendation System Based on Data Mining”, IEEE International Symposium on Electronic Commerce and Security, 2008.

[15] Cheng Qiao, Huang Jian, “Simulation Resource Recommendation System Based on Collaborative Filtering”, Chinese Automation Congress (CAC), 2013.

[16] Qifeng Yang, Sihang Zhang, Bin Feng, “Research on Personalized Recommendation System of Scientific and Technological Periodical Based on Automatic Summarization”, First IEEE Symposium on Information Technologies and Applications in Education, 2007.

[17] BingueGui, Xin Chen, “An Online Book Recommendation System Based on Web Service”, Sixth IEEE International Conference on Fuzzy Systems and Knowledge Discovery, 2009.

[18] Szomszor, M., Cattuto, C., Alani, H., O’hara, K., Baldassarri, A., Loreto, V., Servedio, V.D.P.: Folksonomies, the Semantic Web, and Movie Recommendation. In: Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web 2.0 at the 4th ESWC (2007).

[19] S. Nivattanakul, “Access to knowledge based-on an ontology model,” Ph.D. thesis, University of La Rochelle, 2008.

[20] P. Sittichoke and N. Siranee, “Information retrieval system using N-Gram technique,” in Proc. 5th National Conf. Computing and Information Technology, Bangkok, Thailand, 2009, pp. 307-312.

[21] G. Salton, Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer, USA: Addison-Wesley Longman, 1989, pp. 189-295.