PERSONALISED CONTEXT AWARE CONTENT RELEVANT DISEASE PREDICTION AND DIET RECOMMENDATION SYSTEM

T Ramakrishnudu, Tangedipalli Balakrishna
Dept. of Computer Science and Engineering
National Institute of Technology
Warangal, TS, India
trk@nitw.ac.in

Abstract

predicting the disease is plays an important role in improving the public health. The problem of predicting possible diseases reduce some of the diseases may come in the future. The health recommendation system predicts the disease and recommends the suitable diet and exercises. Context aware recommendation systems produce more relevant recommendations with the help of contextual information. In this paper we have proposed a context aware recommendation system to predict diseases based on the context of the user and recommend a suitable diet and exercises. The experimental results show that the performance of our proposed system is an efficient in predicting the disease and recommending the diet and exercise.

Keywords: Context aware, recommendation system, Similarity Function, prediction, health, diet recommendation.

1.0 Introduction

A recommendation attempts to narrow down selections for users based on their expressed preferences, past behavior, or other required data. Different recommendation systems have been extensively studied in the literature first introduced it in (Otebolaku, A. M. et al, 2011; Vincent Wenchen Zheng et al, 2010) and several types of recommendation systems have been proposed in (Meng, Shunmei, et al. 2014, Chen Yan-Ying et al., 2013, Chen Li et al., 2013) and several context-aware recommendation systems like for restaurant, tourist based and location based studied in (Kyoung Jae Kim et al, 2010; Yu Zheng, et al, 2009 ; Lin Y.-F, et al, 2014).

The recommendation systems are classified based on their nature of working. Basically the recommendation systems are classified into four classes as Collaborative filtering Approach, Content based recommender systems, Hybrid recommendation systems, and Location based recommendation systems.

Nowadays people can easily get their health status like glucose, blood pressure and other parameters. Self-management is the key factor in preventing major disease, but only few people can keep the healthy living style with an enough exercise and good diet habits. Most of are getting the major diseases younger than ever, because of their busy and high pressure modern life style.

Some of the health reports show that the major parts of the causes of the deaths are chronic diseases. Major causes of death are some kinds of chronic diseases, such as the heart diseases, hypertensions, diabetes, chronic liver diseases or kidney diseases. In the high-tension modern life, the result seems reasonable, but that it should not be acceptable, because there are many self-health management ways to make people dodge from the chronic diseases. Unlike cancers, most chronic diseases can be well controlled, even can be prevented, if people can live with a healthy lifestyle, such as the enough sleep, balanced diet and proper exercise.

The use of context-information for predicting diseases and recommendations is a good idea to improve performance of the recommendation system. In this paper we have proposed a context-aware health system to predict diseases and recommendation a diet.

The rest of the paper is organized as follows. Section 2 briefly presents the relevant concepts and definitions. In section 3, existing recommendation systems and challenges are presented. The problem description and proposed method is presented in Section 4. Section 5, illustrate the experimental results. The concluding remarks are finally made in Section 6.

2.0 Concepts and Definitions

The Similarity is one of the fundamental concept in recommendation systems and measured by how much closer two items are related. It is used for determining the degree of matching between the query and item. Similarity score is a normalized value ranges from 0 to 1. There are several functions existing for finding out the similarity value.

Definition1: The similarity between two numeric contextual preferences is calculated as given below:

$$Similarity(Cn, Ce) = 1 - \frac{|C_n - C_e|}{Max - Min}$$

$$Total\ Similarity\ (Cn, Ce) = Similarity\ (Cn, Ce) * Weight$$

Here, Cn, is the new context preference, Ce is the existing context preference, Min and Max are the maximum and minimum value of the ith contextual preference respectively.

Definition2: Precision of disease X is given below:

$$\Pr ecision(X) = \frac{\sum OccurrencesofDiseaseX}{TotalWeight}$$

Here, Total Weight is the normalized total weights of all the users whose normalized weight is greater than user defined threshold value.

2.1 Assigning Weights to Context Variables

When comparing two contexts, the variables which are used to compose the context are to be compared. In principle the entire context variable can be considered as equally important. All the context variables are not equally important, some of the context variables are more important than the other variables. Hence, user defined different weights are assigned to different variables. In this paper, the weights [0 to 1] are assigned to the context variables; the variables which are more relevant to the domain will get more weight i.e one and the variables which are not relevant to the domain will get less weight i.e zero.

2.2 Data Normalization

Data normalization is done in order to find the measure of exactness it determines the fraction of relevant items retrieved out of all items retrieved. We can understand very easily that if the sum of contexts total similarity is greater than the user defined threshold value then it initially computes the occurrences of each context variable and then divides it with the sum of all the context variables. To obtain the precision of each individual context variable after the normalization if more number of the diseases remains then there is a threshold value even for normalization. If the precision of each disease is more than that, then only it includes the diseases to the predicted diseases list.

3.0 Related Work

Jin, Yohan, et al. (2010) proposed a method which includes item-to-item collaborative filtering to discover meaningful interesting videos among the large scale of the videos. Meng, Shunmei, et al. (2014), proposed a method, which presents a personalized service recommendation list and recommending the most appropriate services to the users effectively. In this they used the collaborative filtering approach.

Chen Yan-Ying et al., (2013) proposed an approach, which concentrates on personalized travel recommendation and illustrates auspicious applications by utilizing the freely accessible community-contributed photos. Chen Li et al., (2013), proposed a novel clustering approach built on Latent Class Regression model (LCRM), which is basically ready to consider both the general ratings and feature-level opinion values to observe reviewers' inclination homogeneity.

Lin Y-F, et al., (2014) presented a social media-based recommender system which makes recommendations by considering a user's own health concerns, the trustworthiness of the information providers, the similarity between the user and the information provider, and the test item's general acceptance in the social media. In particular, they proposed a semantics-enhanced fuzzy-based model to facilitate recommendation. The model consists of three important factors affecting recommendation in health care social networking environments: trust, similarity and review. The Fuzzy logic is used in the model because it is tolerant of imprecisely defined data and can model non-linear functions of arbitrary complexity. Most importantly, fuzzy logic can accommodate vagueness, intuitive and experiences in modelling recommendation in a healthcare social network, because human observation forms the basis of recommendation assessments. Semantics-based profile similarity metric is adopted to measure the similarity and proposed Personalized Healthcare Recommender Based on Social Media in which item ratings are used in the Similarity function to give the suggestions to the users interactively.

Hengshu Zhu et al. (2014) illustrated how to extract personal context-aware preferences from the context-rich device logs, or context logs for short. It also, exploits the identified preferences for building personalized context aware recommendation systems. In this they presented three effective methods for mining common context-aware preferences based on two different assumptions about context data dependency. If context data are assumed to be conditionally independent, they proposed an algorithm to mine common context aware preferences through topic models. Otherwise, the context data are assumed to be dependent, exploit the constraint-based Matrix Factorization techniques for mining common context-aware preferences and only consider those contexts that are relevant to content usage for reducing the computation complexity, but it takes large amount of time and memory which is not an efficient when there is lot of missing information.

Jerry C.C. Tseng et al. (2014), proposed a system, which analyzes the result of regular physical examination to evaluate the health risk and provide personalized healthcare services for users in terms of diet and exercise guideline recommendations. They developed some interactive ways for users to get the feedback. Also, they developed a system to get the suggestions for health management from the system

Jong-Hun Kim et al., (2009) proposed a personalization diet recommended service for the users who require the prevention and management for coronary heart disease. This service consists of a collection of modules that draws nutrients, which are to be adopted by users based on the collection of some constraints in users, a configuration module that determines the preference of foods through the input of the diet of users, and a scoring module that makes a score for the extractable diet.

4.0 Problem Description and Proposed System

Problem Statement: Similarity is for retrieving the most similar contextual situations and preferences. This is an important phase of the disease prediction process. To determine the expected preferences of the user in the current contextual situation, it requires a method or an algorithm for determining the degree of similarity between contextual situations.

4.1 Proposed System

The proposed system shown in figure 1 which consists three phases: (i) Pre-Processing System. (ii) Prediction System and (iii) Recommendation System.

Pre-processing System: The pre-processing system consists of new user registration, retrieving the context data, finding the minimum and maximum contextual preferences and finding the similarity. In the pre-processing, the registered user enters into the system through credentials. This phase collects the health data such as blood pressure, BMI, temperature, etc. in addition it also collects the symptoms and the highly influenced variable which consists the data about the history of diseases. This

Context log is maintained in the database. If the user is an existing user then it retrieves the existing contextual data. Find minimum and maximum value of each context variable. User defined weights are assigned to each of the contextual variable as described in the section 2.1. Based on the Definition 1, compute the similarity score with respect to each user context log in the database. Further, compute the total similarity score of each user log by multiplying each similarity score with weights. The disease is not depending on the weight of a single context variable; hence calculate the sum of weights of each user. If the sum of each context log is greater than the user given threshold value then it is consider for further processing, otherwise discard context log.

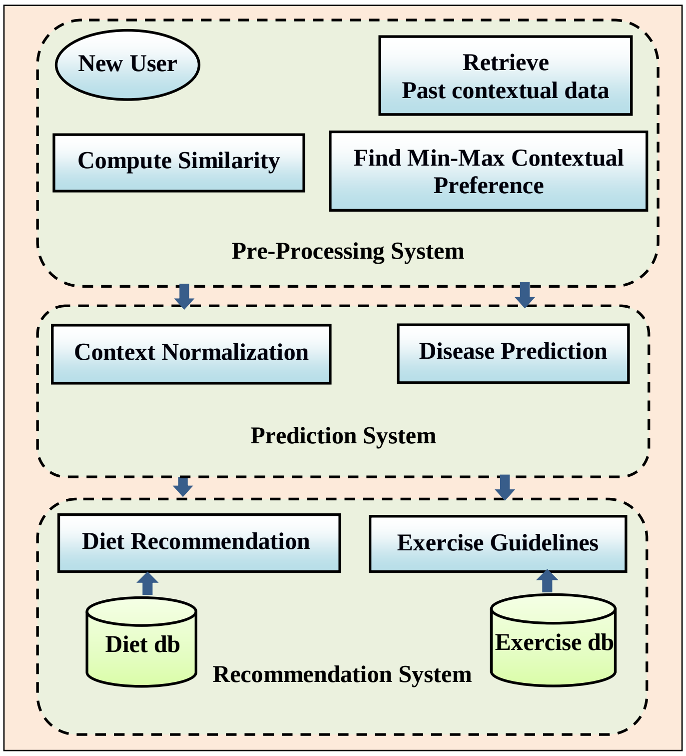

Figure 1. Proposed System

Prediction System: The prediction system consists of two major components one is the context normalization and the disease prediction. In the context normalization each of the disease is considered for the normalization, if it is passed the threshold test. The normalization of each context is done by dividing with sum of all other contexts contribution. Normalization threshold also has to be set to eliminate less probable diseases. In the disease prediction, calculate the precision of each disease in the dataset, if the precision of each disease is greater than the threshold value then those diseases is considered.

Recommendation System: The recommendation system consists of two major recommendations one is the diet recommendation and the other one is the exercise recommendations. For each of the disease a database of diet's and exercises is maintained, which is collected from the expert doctors. After the disease is predicted in the previous phase the diet is recommended based on the disease. Similarly, an exercise is recommended based on the disease.

4.2 Algorithms

Algorithm 1:

Input: New user C_n, Users Database Context Variable list (CV_l), Weight list (W_l), User defined threshold (t)

Output: Similarity Table (ST)

```
Begin
    TS ← 0   /* Ts Total Similarity */
    S ← 0   /* S Similarity */
    for each CV    /* CV: Context Variable  */
      max ← Find max (CV)
      min ← Find min (CV)
    For each Cei   where  i=1 to n
       If  (Cn!=Ce) then
          S = Similarity (Cn,Cei)
          ST ← update(ST,S)
       End if
    End for
    return ST
  End
```

Algorithm 2:

Input: Similarity Table (ST), Weight list (W_l), User defined threshold (t_2)

Output: Diseases (D)

```
1: Begin
2:    Read ST
```

```
For each CV_i in ST
  For each user U_j
     ST←ST(U_j(CV_i)*w_i)  /*Normalization*/
  End for
 End for
 For each user U_j
   For each CV
     Z=Z+U_j(CV_i)
   End for
     If(Z>t_2) then
   sum←sum+Z
  End for
  prec(d_i) = d_i/sum
  if (prec(d_i)>t_1) then
  D =D U d_i
return D
End
```

Algorithm 3:

Input: Diseases (D), Diet and Exercise Database (de_db),

Output: Diet Recommendation (DR), Exercise Recommendation (ER)

```
Begin
   Read user U_i
    If (d_i=d^1)  then /* d_i  is predicted and d^1 in the database */
     diet← diet U d^1. diet
     exer← diet U d^1. exer
    End if
    Rec_list← Rec_list U {diet, exer }
  End
```

5.0 Experimental Results

In this section we measure the accuracy of our proposed method. To evaluate the performance our proposed method we have conducted several experiments. All the experiments conducted in Intel 64bit PC with preinstalled Ubuntu, XAMPP, MySQL, and Python 3.6.2.

5.1 Input Dataset

To evaluate our proposed method, initially we used a synthetic dataset with 1000 users and 20 different diseases. The results are showed in Fig. 2, Fig. 3 and Fig.4. It is a medical dataset, containing details of the patients.

5.2 Results Analysis

We test our proposed algorithm to predict the disease. Figure 2, depicts the accuracy of the proposed method. The proposed method is better in predicting the diseases. Our method shows that the precision of the disease thyroid is approximately 0.78, because the dataset having the enough context log data about the thyroid. Whereas the prediction accuracy of leg pain is 0.54 is very less, because the dataset consists of less context about the leg pain and the user given weightage is plays vital role in the perdition accuracy of the disease.

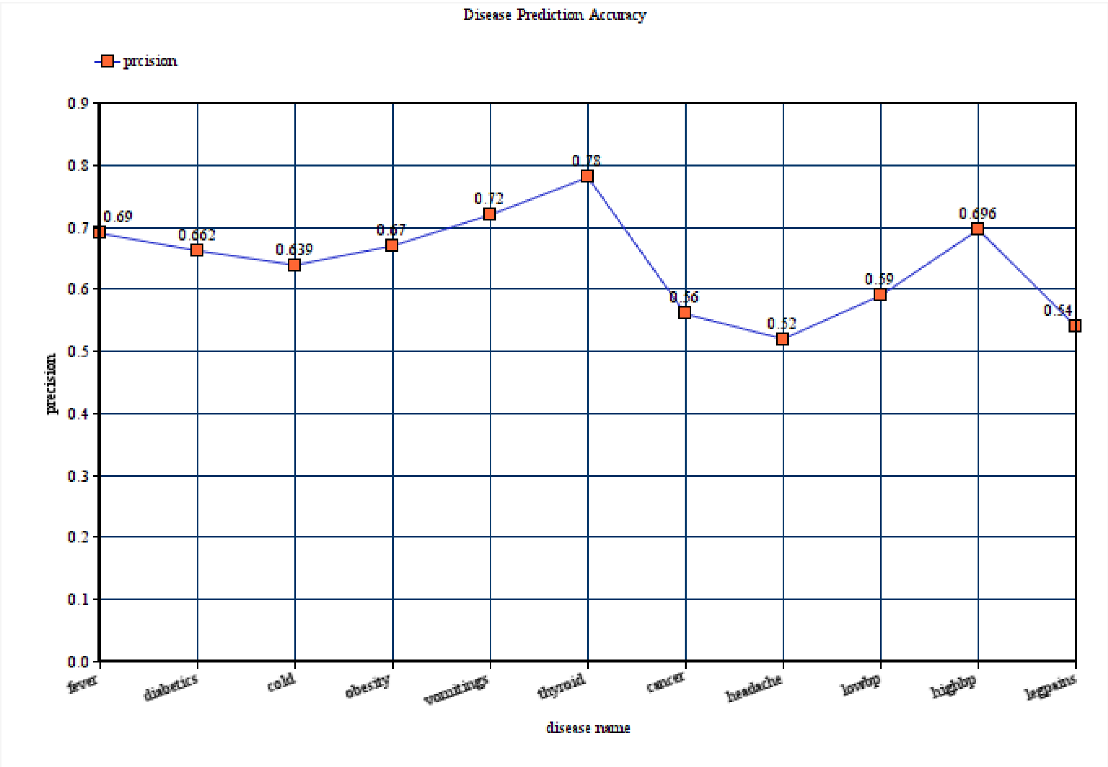

Figure 2: Disease Prediction Accuracy

In this, paper we have proposed our own similarity function. We tested our approach with the proposed similarity function and the Pearson correlation coefficient function. The results are shown in figure 3. When the dataset size is less both the functions gives the approximate similarity score, when the dataset size is more our proposed similarity function is giving a better similarity score, whereas the Pearson correlation coefficient is giving very less score.

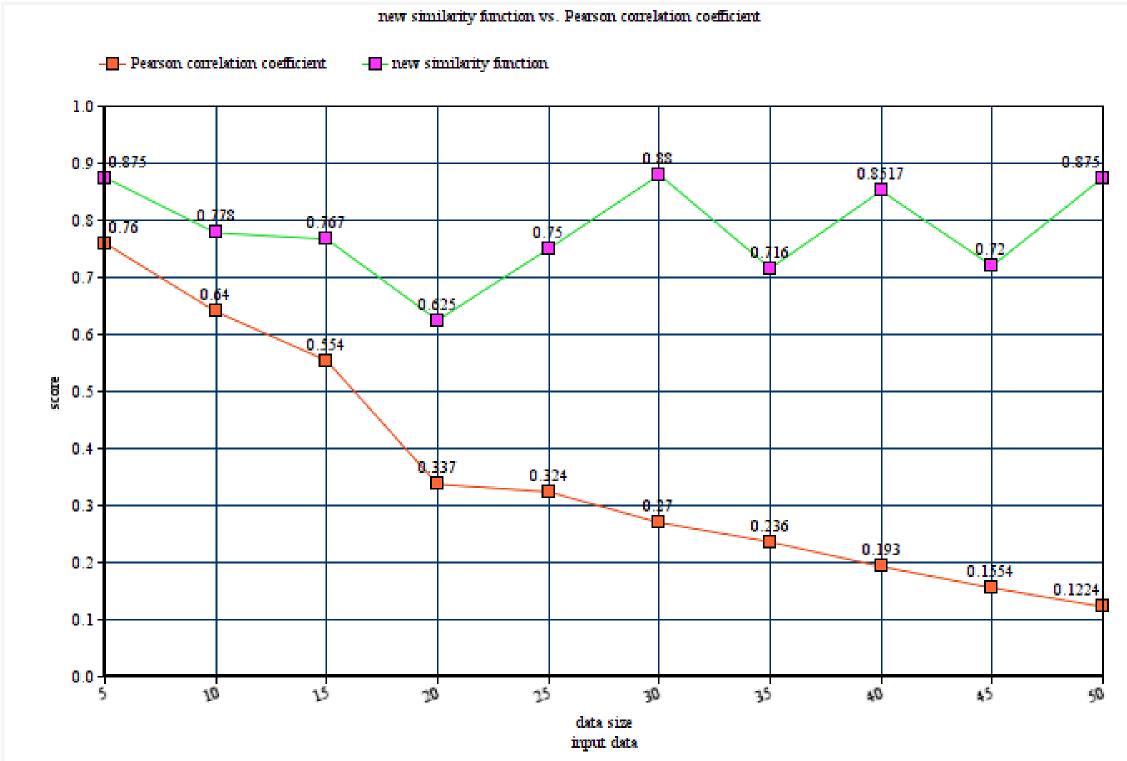

Figure 3: Proposed similarity function vs. Pearson correlation coefficient

Figure 4, shows the comparison results of different recommendation systems. In this we have used several diseases on X-axis and its respective accuracy on Y-axis. Here, we compared our proposed system with the existing Case-based recommendation system and the content-based recommendation system. Our recommendation system gives 76% more accurate results than the other two recommendation systems.

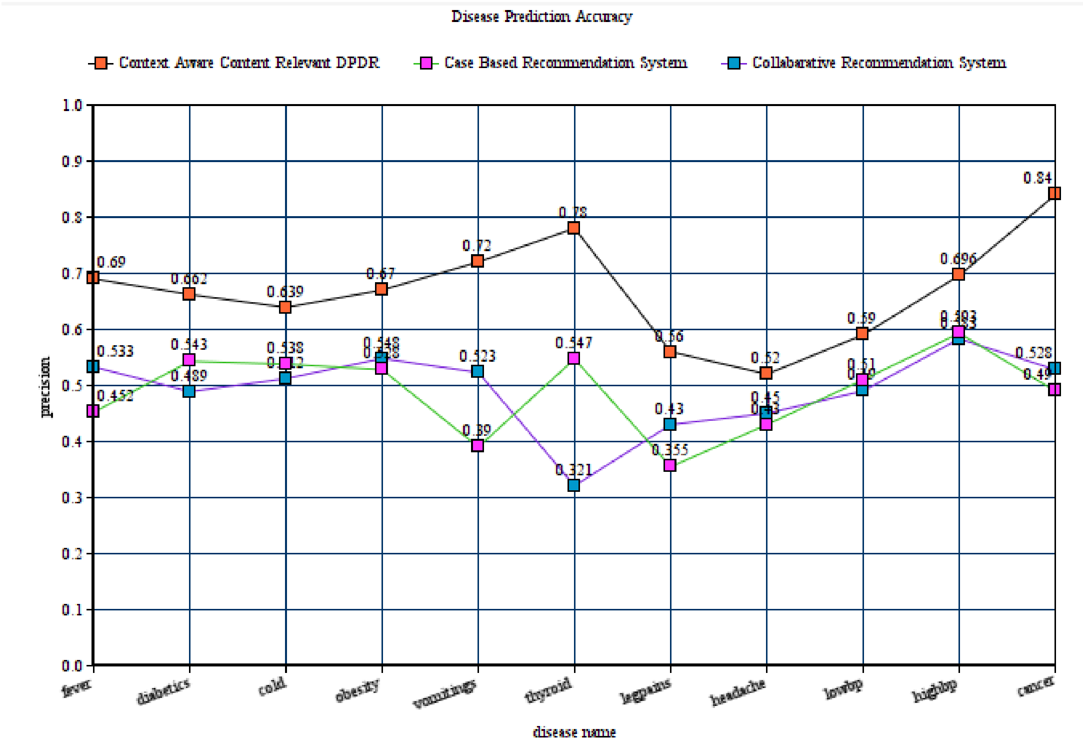

Figure 4: Comparison results between three systems

6.0 Conclusion and Future Work

In this paper, we have proposed a similarity function, which is used to find the similarities between the existing user's context logs with the new user context log. We also, designed a Personalized Context Aware Content Relevant Disease prediction and Diet Recommendation System based the proposed similarity function. The proposed recommendation system based on the new similarity function is gives better results than the existing recommendation systems. The normalization process used in the proposed recommendation system helps to get the most likely diseases. The recommendation system gives better results when the disease context data is available in the context log of the existing users in the database. The diseases predicted in the from the existing log helps to recommend the diet and the exercises.

Our future research work include the design of better similarity functions, which are used to suitable for disease prediction and design of a real time prediction system which helps to predict the disease in real time.

References

Chen, Li, and Feng Wang. (2013) *Preference-based clustering reviews for augmenting e-commerce recommendation,* Knowledge-Based Systems 50 44-59.[13]

Chen, Yan-Ying, An-Jung Cheng, and Winston H. Hsu.(2013) *Travel recommendation by mining people attributes and travel group types from community-contributed photos,* IEEE Transactions on Multimedia 15.6, 1283-1295. [12]

Hengshu Zhu and Enhong Chen.(2014) *Mining Mobile User Preferences for Personalized Context-Aware Recommendation*, ACM Transactions on Intelligent Systems and Technology, New York, Vol. 5, No. 4, Article 58. [1]

Jerry, C.C. Tseng, Kuen-Rong Lo.(2014) *An Interactive Healthcare System with Personalized Diet and Exercise Guideline Recommendation* International Journal of Environmental Research and Public Health (IJERPH), 11(4), pp.4262-4279. [7]

Jin, Yohan, et al. (2010) *MySpace video recommendation with map-reduce on qizmt* 2010 IEEE Fourth International Conference on Semantic Computing (ICSC), Pennsylvania, USA, pp.126-133. [10]

Jong-Hun Kim et al., (2009) *Design of Diet Recommendation System for Healthcare Service Based on User Information*, Fourth International Conference on Computer Sciences and Convergence Information Technology(ICCIT), Seoul, South Korea, pp. 516-518.[14]

Kyoung Jae Kim, Hyunchul Ahn, and Sangwon Jeong. (2010) *Context-aware recommender systems using data mining techniques,* Journal of World Academy of Science, Engineering and Technology, 64, pp.357–362. [5]

Lin, Y.-F., et al. (2014) *A Framework for Personalized Diet and Exercise Guideline Recommendation,* the Conference on Technologies and Applications of Artificial Intelligence (TAAI), pp. 274-283. [9]

Meng, Shunmei, et al. (2014) *KASR: A Keyword-Aware Service Recommendation method on MapReduce for big data applications,* IEEE Transactions on Parallel and Distributed Systems 25.12, 3221-3231. [11]

Otebolaku, A. M. and Andrade, M.T. (2011) *Context Representation for Context-Aware Mobile Multimedia Recommendation,* In Proceedings of the 15th IASTED International Conference on Internet and Multimedia Systems and Applications, Washington.[3]

Roh, J.H. and Jin, S. (2012) *Personalized advertisement recommendation system based on user profile in the smart phone,* In Advanced Communication Technology (ICACT), 14th International Conference, Beijing, pp.1300-1303.[2]

Vincent Wenchen Zheng, Bin Cao, Yu Zheng, Xing Xie, and Qiang Yang (2010) *Collaborative filtering meets mobile recommendation: A user-centered approach,* In Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10). Chile, pp.236–241.[4]

Wu, Chun-Hui, Fang, Kwot-Ting, and Chen, Ta-Cheng (2009) *Applying data mining for prostate cancer*, in the International Conference on New Trends in Information and Service Science, Beijing, pp. 1063-1065. [8]

Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma (2009) *Mining interesting locations and travel sequences from GPS trajectories,* In Proceedings of the

18th International Conference on World Wide Web (WWW’09) ACM, New York.[6]