Volume 6, Issue 4, April 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Food Recommendation Systems

Shashank S U
Information Science Dept.
National Institute of Engineering,
Mysuru, India

Raksha Hegde M
Information Science Dept.
National Institute of Engineering,
Mysuru, India

K Raghuveer
Professor and Head
Information Science Dept.
NIE, Mysuru, India

Abstract— This paper aims to give the survey on food recommendation system along with a proposed food recommendation system. This recommender uses content based technique in order to give food recommendations. Keyword similarity is found using a hybrid similarity measure.

Keywords— Food recommendation system; content based technique; hybrid similarity measure; recommender systems

I. INTRODUCTION

In our day to day life we come across questions like "which restaurant to have lunch?" "Where do we spend our evening?" etc. As we know every user want a list of options from which he/she can choose the best one. Recommendation is an act of saying something is good and deserve to be chosen. The recommendation system provides a wide range of options to the users to choose one of the best options among them.

We, nowadays find a lot of restaurant recommendation system which are based on rating. In this paper, we are coming up with an idea of recommending food items of user's interest along with the restaurant. Recommendations can be using different techniques like content based technique, Collaborative filtering based etc. This paper uses content based recommendation for food recommendation.

II. EXISTING SYSTEM

In menu recommendation system, Menu recommends available menus based on user preference. System provides a unified platform to display menus of various canteens, Users can filter meals according to their taste and rate them for latecomers to benefit from their fellow colleagues' review [4]. Good food restaurant recommendation system [1] combines the ideas from different blogs. It improves k-means algorithm using white ants habit model for better clustering performance based on evaluation and comparison of existing k means. Within the cluster groups, paper suggests a recommendation system that provides rank based on similar levels in accordance with certain topics.

By using content based recommendation system, food is recommended based on item description as well as profile of the user's interest in [2]. This system involves user profiling, food profiling and recommending foods based on previous feedback of user. Initially after registration user is asked to rate a set of random ingredients. Based on rating similar food is calculated and recommended to user. Similarity of items are found using method called Inverse document frequency. User feedback is taken for the recommendation made and next recommendation is built on both personally liked ingredients and the previously liked recommendations. Dataset which they have used is from "yemeksepeti.com" and project is implemented in php. [3] aims to develop a predictive model of customers' restaurant ratings using principles and techniques of recommendation system. They have made use of yelp dataset to extract collaborative and content based features to identify customer and restaurant profiles. Project focuses on creating a recommendation system for yelp users to potential food choices they could make. They have implemented hybrid cascade of K-nearest neighbor clustering, weighted bi-partite graph projection and other learning algorithms. Algorithms performance are evaluated using Root metrics mean squared error and Mean absolute error.

III. RELATED THEORIES AND STUDIES

A. *Keyword Extraction*

In Simple approach we compute similarity of an unseen item based on the keyword overlap.

$$\sin(b_i, b_j) = \frac{2 * |keywords(b_i) \cap keywords(b_j)|}{|keywords(b_i)| + |keywords(b_j)|}$$

Disadvantages of this method are that not every word has similar importance and longer documents have a higher chance to have an overlap.

Another measure that we can make use of is standard measure: TF- IDF. This measure represents text documents as weighted term vector. TF (Term Frequency) measures, how often a term appears (density in a document). IDF (Inverse Document Frequency) aims to reduce the weight of terms that appear in all documents.

Here N denotes for the total number of foods in a collection, and n (i) is for the total number of foods that have the ingredient "i". Thus the IDF of a rare term is high, whereas the IDF of a frequent term is likely to be low. Hence IDF will be useful in order to find out frequent ingredient in a list.

B. Keyword Similarity Measures:

1) Character Based Similarity Measures

a) Jaccard Similarity: Jaccard Similarity Coefficient is used to compare characteristic similarity between sets of documents.

$$Jaccard = \frac{|A \cap B|}{|A \cup B|}$$

A = set of characters that doesn't occur in word 1
B = set of characters that doesn't occur in word 2

b) N-Gram similarity measure: N-Gram similarity measure is used to calculate probability of word sequence that occurs as a sentence. Probability of word sequence can be estimated from source of data. N-Gram size can be from 1 to n.

$$Ngram(2,X) = \{x_0, x_1, x_2, x_3, x_4 \dots \dots \dots \dots \dots x_{n-1}, x_n\}$$

c) Vector space model: Vector space model or term vector model is an algebraic model for representing text documents (and any objects, in general) as vectors of identifiers

$$Vectorspace = \left(\frac{V_A}{len(A)}\right)\left(\frac{V_B}{len(B)}\right)$$

V_A=Vector of number of characters in word 1
V_B=Vector of number of characters in word 2
A = set of all characters in word 1
B = set of all characters in word 2

d) Longest Common Substring (LCS): LCS calculates similarity based on the length of contiguous chain of characters that exist in both strings.[7]

e) Damerau-Levenshtein: This finds the similarity by counting the minimum number of operations needed to transform one string into the other, where an operation is defined as an insertion, deletion, or substitution of a single character, or a transposition of two adjacent characters.[7]

f) Jaro: This finds the similarity based on the number and order of the common character between two strings. Typical spelling mistakes are taken into account in this method.[7]

g) Jaro–Winkler: Jaro–Winkler is an extension of Jaro distance; it uses a prefix scale which gives more favorable ratings to strings that match from the beginning for a set prefix length.[7]

2) Corpus Based Similarity Measure: This similarity measure is a similarity measure that determines the similarity between words according to information gained from large corpora. A Corpus is a large collection of written or spoken texts that is used for language research.[7]

3) Knowledge-Based Similarity Measure: This Knowledge Based Similarity is one of semantic similarity measures that bases on identifying the degree of similarity between words using information derived from semantic networks [8]. WordNet [9] is the most popular semantic network in the area of measuring the Knowledge-Based similarity between words similarity measure is a similarity measure that determines the similarity between words according to information

4) Hybrid Measure: [6] has introduced Jaccard's, N-Gram,
Vector space, Average (JNVA) and Jaccard's, N-Gram, Length, Average (JNLA) by using hybrid method.

$$JNLA = \frac{jaccard + Ngram + hlengt}{3}$$

$$hlengt = exp\left[\frac{-abs(len(A) - len(B))}{len(A \cup B)}\right]$$

$$JNVA = \frac{Jaccard + Ngram + Vectorspace}{3}$$

C. Document similarity measure

Similarity based on keywords or phrases can be calculated using this function:

$$Similarity_{Keyword}(p,q) = \frac{KN(p,q)}{Max\left(kn(p), kn(q)\right)}$$

Where *kn*(*p*), *kn*(*q*) is the number of keywords in a document, *KN*(*p*, *q*) is the number of common keywords in two documents.

$$s(d_i, d_j) = \frac{L\left(F(d_i, d_j)\right) - 1}{L_{Total} - 1}$$

IV. PROPOSED SYSTEM

Every day we form opinions about things we like, don't like and even don't care about. Recommendation is all about predicting these likes and dislikes, and using them to discover new and desirable things you didn't already know about.

First we will scrape food items and store it in a database and then we will scrape ingredients and store it in a database. Then using idf we aim to reduce the weight of the ingredients that appear in all documents. Now we will find similarity measures between two documents (one is document with food items and other documents with ingredients obtained after applying idf). It is done using similarity based on keywords (JNLA). We take each ingredient obtained by applying idf as keyword.

For each food item in the database, the ingredients are listed and it is compared with the keywords using similarity based on keywords. We will fix a threshold and if the similarity measure is greater than the fixed threshold we will add that food item into a cluster. In this way we will make a cluster of food items based on the user input. Food items in the cluster are recommended for the user.

V. ALGORITHM

1) Algorithm

Input location and food
Scrape the ingredients from the web
Create a list with the importance of keywords using idf
Call clustering method for clustered database
Output recommended food items and recommended restaurants at that location or near to that location

2) Clustering method

```
for every keyword_set in cluster_keyword_set
    for every keyword in keyword set
        for every keyword in Cluster_keyword_set:
            sim_keyword = JNLA( keyword, doc_word)
            if(sim_keyword > threshold)
                match_count++
        end for
    end for
    similarity[cluster_no] = match_count / max( no_keywords in document i, no_keywords in keyword set)
end for
MaxSimCluster = max(similarity[])
If ( NoOfClusters in MaxSimCluster > 10)
Go to Step 2 for MaxSimCluster
Else
Output the cluster
```

REFERENCES

[1] E-Seok Jo, Yong-Sung Kim ,"A Good Food Restaurant Recommendation System Using Termite", International Journal of Advancements in Computing Technology(IJACT), Volume 5, Number 12, August 2013
[2] Ipek TATLI , "Food Recommendation System", June,2009
[3] Sumedh Sawant, Gina Pai, "Yelp Food Recommendation System"
[4] Marian Runo, "FooDroid: A Food Recommendation App for University Canteens", June 21, 2011
[5] Seung-Shik Kang, "Keyword-based Document Clustering"
[6] Jatsada Singthongchai and Suphakit NiwattanakulK. Elissa, "A Method for Measuring Keywords Similarity by Applying Jaccard's, N-Gram and Vector Space", Lecture Notes on Information Theory Vol. 1, No. 4, December 2013
[7] Wael H. Gomaa, Aly A. Fahmy, "A Survey of Text Similarity Approaches", International Journal of Computer Applications (0975 – 8887) Volume 68- No.13, April 2013
[8] Mihalcea., Corley, & Strapparava, "Corpus based and knowledge-based measures of text semantic similarity", Proceedings of the American Association for Artificial Intelligence 2006
[9] Miller, Beckwith, Fellbaum, Gross & Miller, "WordNet: An online lexical database"., Int. J. Lexicograph. 3, 4, pp. 235–244, 1990

