

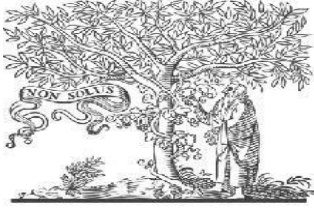


International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2019IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 1st Jun 2019. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-06](http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-06)

Title: **RECOMMENDATION SYSTEM OF FOOD PRODUCTS BASED ON NUTRITIONAL VALUE**

Volume 08, Issue 06, Pages: 27–32.

Paper Authors

NEHA M L, YAGYA DEV, Y.VAMSI KRISHNA

Sri Krishna Institute of Technology, Bangalore.



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

RECOMMENDATION SYSTEM OF FOOD PRODUCTS BASED ON NUTRITIONAL VALUE

¹NEHA M L, ²YAGYA DEV, ³Y.VAMSI KRISHNA

¹8th sem Dept.of ISE,Sri Krishna Institute of Technology,Bengaluru , Karnataka, India

²8th sem Dept.of ISE,Sri Krishna Institute of Technology, Bengaluru, Karnataka, India

³Prof Dept.of ISE,Sri Krishna Institute of Technology,Bengaluru, Karnataka, India

¹Neharao3297@gmail.com,²Yayadevsingh24@gmail.com,³Yvnmk.1229@gmail.com

Abstract— Monitoring of food in take in our daily life is an important task. The food recommender system aims to assist the user in daily diet selections based on some nutrition guidelines. In this paper we propose a system where it recommends the food product based on the nutrition value and develops a genetic process where it combines the customer nutrition and the food nutrition. Here Deep learning neural network model is applied to automatic product categorization. This model is better when compared to SVM, Naives bayes, and KNN model.

Keywords—Recommendation system, person-alized diets, deep learning, genetic algorithm.

I Introduction

For the better production quality and diet choices Expert systems have been expected as a solution since 1998 [1].System that can cope up with variety volume and velocity use Big data technology to bring success to business. [2], science and health-care[3]. Product recommendation took birth when many recommendation systems aimed to scaled with Big data through MAP-Reduce [4]. Ease of use [5].In this paper, we Combine existing algorithm to construct a application for food products recommendation. As shown in Fig. 1, the included algorithms cover the task of data categorization, data analysis and decision recommendation.

In any given data there will be number of noises present. In order to deal with the noises the Deep Neural Network(DNN) is used which have outstanding capability in dealing with data containing noise. Optimization in Nutri-Flow Software is needed to find a level of change for a set of foodstuffs to reach a more balanced diet.All the molecules involved in human metabolism are directly or indirectly controlled by the genes. And by this individual health can be optimized through their personalized health profile. In this application we will generate a health profile for the user and we will also categorize the food products based on the few features and the Genetic algorithm where it combines the user health and the categorized food products and build a relationship between them. The proposed expert recommendation system framework consists of four main components:

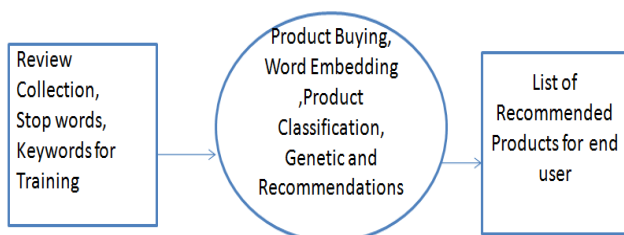


Fig: Overview Of The Recommendation System

- 1) Data extraction is done using web crawler. Where it collects the data either in online or offline mode.
- 2) Word Embedding model is introduced for converting text data into generalised vectors.
- 3) A DNN model for product categorization based on few features and this model can perform with good accuracy.
- 4) Recommendation products where it recommends the food products for the user based on nutrition value which includes Genetic algorithm.

II. WEB CRAWLER

The Web Crawler based Online Review Submission will have the input elements like product, Web URL and Xpath. The Web URL and Xpath will be validated. If valid then a real HTTP connection is made to the web site and the DOM elements are scanned for text (review) and extracted.

Web Crawler based Review Collection Process

1. Web URL, Xpath and Product will acts like an input
2. The HTTP Connection is made to the Web URL
3. The Complete HTML of the Web Site is Downloaded
4. The DOM tree is constructed
5. The Xpath is used and the corresponding Node of the DOM is navigated.
6. The count number of similar DOM Nodes is found out
7. Starting from the 1st DOM node till the count of DOM Nodes
 - a. Obtain the text of the DOM Node which satisfies the Xpath
8. The Review is Stored in format {ReviewID, ProductID, Review Desc}

III. DEEP NEURAL NETWORK FOR PRODUCT CATEGORISATION

WORD EMBEDDING

This algorithm first takes the training data. Finds out the unique words from the training data, computes the occurrence of the words and then generates a matrix of N*M word vector where N is number of reviews and M is the number of unique words across all the reviews. This is very cumbersome process. Hence a cron job will be running which runs throughout the day and computes it. Once the word vector is found out then it acts as an input for the classification the hidden neural nodes establishes the relationship between various nutrition factors and produces a labeled output for the products. The label is classes like sugar, fat, energy etc.

The Word Embedding Algorithm can be described as divided into the following modules

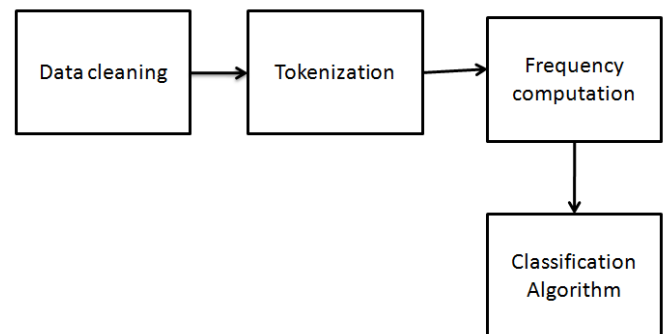


Fig: Sub Modules Of Word Embedding Algorithm

A. Data cleaning

Data Cleaning Algorithm is responsible for cleaning of reviews which have been collected either in offline mode or online mode. The Data Cleaning is performed by making use of stop words and removing the stop words from the reviews.

1) WORD SEGMENTATION

Word segmentation refers to the cutting of coherent vulnerability text information into one word, which transforms the entire vulnerability text information into the smallest semantic unit that can be counted by statistics.

2) WORDS STEMMING

Lemmatization or Words stemming refers to the transformation of non-root form into root form in the word set, and that is the verb in English description. According to the data mining words with similar meaning should be considered as one single word. For example eat, eating, ate should be considered as single word eat.

3) STOP WORD FILTERING

Stop word filtering refers to words that appear frequently in text and contribute little or no contribution to the content or classification of text information; it includes both public stop words and professional stop words.

Few of the Stop words are Listed in the below paragraph

a,about,above,across,after,afterwards,again,against, all,almost,alone,along,already,also,although,always, am,among,amongst,amoungst,amount,an,becoming, been, before, beforehand, behind, being, below and beside

Steps involved in Data Cleaning Algorithm

1. The List of Reviews are obtained
2. The Measure for the count of Number of Reviews is Obtained
3. Starting from the 1st Review till the number of reviews. Repeat the following process
 - a. Obtain the Review Description
 - b. Convert the Review Description into a FIFO Queue by diving the entire sentence into a set of words using a Delimiter
 - c. Measure the Count for Number of words present in the review
4. Start from the 1st word till the number of words. Repeat the following process
 - a. Obtain the Word
 - b. Compare against the list of stop words
 - c. If the word is a stop word then the word is skipped and next word is scanned

- d. If the word is not a stop word then one more FIFO Queue is constructed with word which is not a stop word.
- e. Once all the words have been scanned then the Clean Review is Stored

B. TOKENIZATION

Tokenization is a process of converting the clean data into a set of words known as tokens. Steps involved in Tokenization Algorithm

1. The List of Clean Reviews are obtained
2. The Measure for the count of Number of Reviews is Obtained
3. Starting from the 1st Review till the number of reviews. Repeat the following process
 - a. Obtain the Review Description
 - b. Convert the Review Description into a FIFO Queue by diving the entire sentence into a set of words using a Delimiter
 - c. Measure the Count for Number of words present in the review
4. Start from the 1st word till the number of words. Repeat the following process
 - a. Obtain the Word
 - b. Store the Token in the following Format {Token ID, Token Name, Review ID, ProductID}

C. FREQUENCY COMPUTATION

This is a process in which the frequency computation is performed. For each of the reviews the frequency is computed. Frequency is number of times a i^{th} token appears in j^{th} . Review. Steps for frequency computation

1. The List of Tokens are obtained
2. The Measure for the count of Number of Tokens in Reviews is Obtained
3. Starting from the 1st Token till the number of reviews. Repeat the following process
5. Obtain the Token

6. Convert the Token into a FIFO Queue and count the Token
4. Find the Frequency of the Word
5. Store the Frequency in the following Format
{Freq ID, Token Name, Frequency, Product ID, Review ID}

D. CLASSIFICATION ALGORITHM

1. Find the Unique Products from the reviews
2. Perform the frequency computation only fat, sugar, salt, energy related terms
3. Find the Product Matrix in the following format

Product name	Fat Freq	Salt Freq	Sugar Freq	Energy Freq	Protein Freq	Iron Freq
--------------	----------	-----------	------------	-------------	--------------	-----------

1. Find the maximum value of each row. The product is classified as belonging to that specific category
2. If more than one column has the same frequency and it is maximum then product is assigned for both category

IV. SYSTEM ARCHITECTURE

Fig shows the various end points. As shown in the fig the web.xml allows only valid request into the web container maintained by tomcat. The entire User Interface is designed based on JSP, Ext JS and Angular JS. The REST Ful Services are used for entire algorithm and business services implementation. The data is stored and retrieved in the data store by the respective Service.

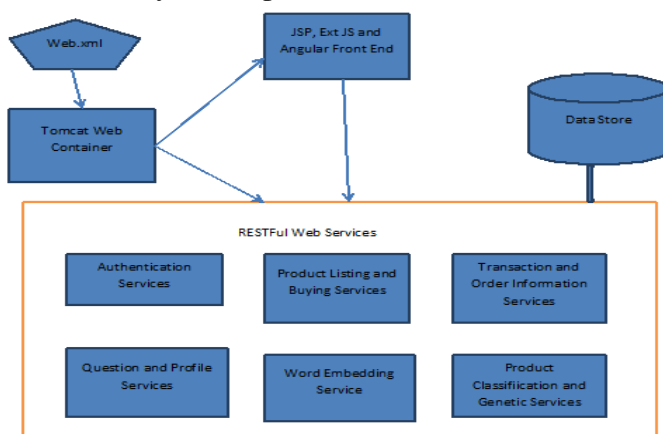


Fig: Optimized Personal Nutrition Service

V. NUTRITION QUESTIONNAIRE

User Profile Generation and Nutrition Questioner

The User Profile is generated based on the following questions

1. User will be asked a set of questions. Each Question will be having the following attributes
 QuesID- An Autogenerated ID for the Question
 QUESDESC- This is the question description
 ANSWER1- This is the Answer1 visible to User
 ANSWER2- This the Answer2 visible to User
 ANSWER3- This the Answer3 visible to User
 ANSWER4- This the Answer4 visible to User
 TESTNAME- This is the Name of the Test
 RATING1- This is the rating for answer1
 RATING2- This is the rating for answer2
 RATING2- This is the rating for answer3
 RATING2- This is the rating for answer4
 TYPE- It can be SUGAR, SALT etc
2. Based on the Answer of the user the rating for the user is generated
3. The Classification of the user profile is performed

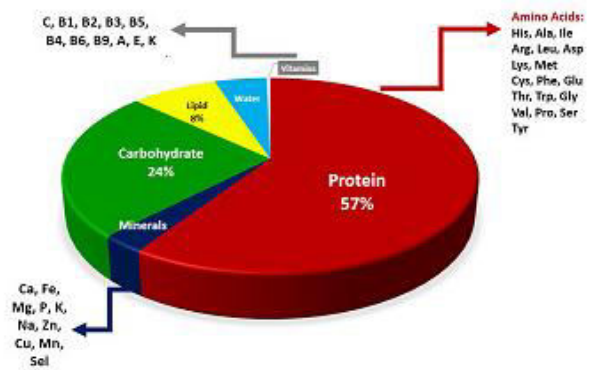


Fig: Nutrition Intake For A Human

VI. GENETIC ALGORITHM

Genetic algorithms are based on processing of a population of coded solution alternatives.

1. Generate Population Set

This phase is responsible for represents each of the product based on the nutrients in terms of the numerical values. The nutrition value now are represented in a binary value this creates a population set like these many sets are generated.

2. Evaluation

This computes the value in a normalized format. It counts the number of zero's and one's in a population set.

3. Selection

The top products are chosen based on questioner biological factors of the user.

4. Crossover

Crossover operators on population towards better solutions by distributing good genetic matter between generations. Crossover starts by selection. It selects the population set and the product features are divided into equal halves.

5. Mutation

This process is responsible for changing one of the value to generate new population set. Here actually all the binary values are considered and the number of one's and zero's are counted and any one of the binary value is changed to one or any zero value is changed to one. So that fro this process a new population set is generated.

VI. COMPARISION AMONG CLASSIFICATION METHODS

Many machine learning algorithm Such as SVM, Naives bayes and KNN have been used for classification but there are few drawbacks in these classification models. SVM is an algorithm that determine the best decision boundary between vectors that belong to a given group and vector that do not belong to. Drawback of this method is that it has several key parameters that need to be set correctly to achieve the best classification for any problem. Naives bayes classifier are linear

classifier. The probabilistic model of naives bayes classifier is based on Bayes theorem and the adjective naïve comes from the assumption that the dataset are mutually independent. Drawback of this method is that in reality any two features are dependent on each other and one more drawback is that data scarcity we take frequency of features but here it takes 0 or 1 getting worse result. KNN algorithm for text classification is used to classify by finding the K nearest matches in training data and then using the label of closest matches (uses Euclidean distance formula) to predict. Drawback of this method is that as KNN requires scaling of data because it uses Euclidean formula and one more disadvantage is that it is slower and costlier in terms of time and memory(large memory for storing training data set).

VII. CONCLUSION

In this paper we proposed a Recommendation System of food products based on nutrition value were it recommends food product based on nutrition value of the user. Here it uses DNN model to classify and categories the products based on features like salt, sugar.fat, energy, protein, iron. Web crawler collects the reviews from online either online or offline and the word embedding algorithm using DNN model classifies the product. And the nutrition questionnaire is performed where the user gives answers to the authenticated questions and thus health profile of the user is generated and now the genetic algorithm comes to picture where it combines or considers the nutrition value of the user and the classified product value and generates a relationship between the user and the product. And thus this product is recommended to the user.

References

- [1] S. Linko, "Expert systems what can they do for the food industry?" *Trends Food Sci. Technol.*, vol. 9, no. 1, pp. 3–12, 1998.
- [2] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton, "Big data,"



- Manage. Revolution Harvard Bus. Rev.*, vol. 90, no. 10, pp. 61–67, 2012.
- [3] M. Hansen, T. Miron-Shatz, A. Lau, and C. Paton, “Big data in science and healthcare: A review of recent literature and perspectives,” *Yearbook Med. Informat.*, vol. 9, pp. 21–26, 2014.
- [4] G. Linden, B. Smith, and J. York, “Amazon.com recommendations: Item-to-item collaborative filtering,” *IEEE Internet Comput.*, vol. 7, no. 1, pp. 76–80, Jan./Feb. 2003.
- [5] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, “Promises and challenges of big data computing in health sciences,” *Big Data Res.*, vol. 2, no. 1, pp. 2–11, 2015.