

IMPROVING PREDICTION THROUGH THE USE OF C-KNN A HYBRID APPROACH OF FILTERING INDICED IN KNN

¹Jaskiran Ghotra, ²Mini Singh Ahuja

¹Student, ²Assistant Professor

^{1,2} Department Of Computer Science, Guru Nanak Dev University Regional Campus, Gurdaspur, (Punjab) India
¹ghotra.jaskiran@gmail.com, ²minianhad@yahoo.co.in

Abstract In today's environment the idea of establishing business without the use of internet is not possible. More and more users are shifted towards online systems. So companies are also converged toward the online business. Every company in their attempt to establish strong roots required some sort of mechanism which can promote their product. So recommender system comes into existence. The recommender system is the filtering system which will detect the preferences of the users. By looking at the preference of the users companies can decide which product to be launched in the market and which is not. So recommender system is the need of the hour. Recommender systems are used for wide variety of applications which includes movies, music, news, life insurance etc.

Keywords Recommender System, Users, Online System, online business

I. INTRODUCTION

In order to increase the sale of the product or to increase the efficiency in the market recommender system is needed. This system will require large amount of information in order to make correct decision. The information which is provided to the recommender system must be consistent in nature. For the information some sort of information system is required. The recommender system will take the information and formulate the decision in one of the following two ways- either by the use of collaborative filtering or by the use of content filtering. The collaborative filtering is the mechanism of filtering for information among the multi agents, viewpoints, data sources etc. The content filtering on the other hand is the mechanism of using the program in order to filter the information which is going to be used within the system. People now days are more and more concerned with the environment. They are conscious of the fact that the product used by them may harm the environment. So people are inclined toward the safety of the environment. So we propose a recommender system which will promote Green Products only. In the proposed system content filtering will be utilized.

Today most of the people consider saving our environment. Our environment is corrupted by large number of gases and other harmful products which are produced by the human beings. So, most of the people consider preserving our environment and hence Green Product. In the proposed system we will construct Recommender System which will be going to promote the products which are environment friendly. Recommender System will help in promoting the Green products.

II. RELATED WORK

The brand awareness is a massive issue which must be considered in the task of promotion. The brands which are popular are likely to be sold. So task of promotion is very important. Promotion of Green Products (Barreda, Bilgihan, Nusair, & Okumus, 2015) is considered so that environment remains clean. The promotional work although is convoluted but is accomplished and counseled in the existing work. There are legions which can inspire the promotional process. These will include system quality and information quality (Barreda et al., 2015). In accession to the above proposition the recommender system is also common in the applications of ecommerce (Thai-Nghe, Drumond, Krohn-Grimberghe, & Schmidt-Thieme, 2010). The recommender system is commonly used in the area of e-learning. It will be used to evaluate the performance of the students. The experimental results shows that the performance of the system improved by the use of recommender system in the area of e-learning (Thai-Nghe et al., 2010). It is considered that the preferences of the user does not change over the period of time and decision can be taken by looking at the historical data (Baltrunas, 2011). This is true to some extent. However preferences of the user may be influenced by legions of other factors also (Baltrunas, 2011). As an example consider a rainy season in which user will not prefer to go to a beach rather user prefer to go to a museum (Baltrunas, 2011). The context sensitive algorithm is used in this case. The hypermedia supported education can also be accomplished using the recommender system (Hoic-Bozic, Holenko Dlab, & Mornar, 2015). The blended learning model is followed in this case. When this happens personal online learning will be enhanced (Hoic-Bozic et al., 2015). Online community of practices is another application of

recommender system. The trust based COPs is created so that online education can be promoted(Zheng et al., 2015). The stronger social relations are created by the use of COPs. The hybrid algorithm will provide more accurate suggestions as compare to content based recommender system(Zheng et al., 2015). By analyzing the background of the recommender systems we conclude that legion of work has been realized in the area of e-learning. Final work has been done in the area of beget of recommender system to promote green products.

III. List of Products Considered for evaluation along with attributes to be satisfied

There are number products and number of companies which are considering safety of the planet and proceeding toward preserving the environment. So people more and more converging towards the utilization of Green Products. In the proposed system the products which we are considering are electronic products such as LED, Microwave, and Air Conditioner etc. In order to decide whether the product is Green or not number of parameters is to be considered.

There exists number of products each having certain characteristics. In order for the product to be Green certain properties are needed to be satisfied. Some properties are required(R) and some properties are optional (O) in nature. We need to identify which product is appearing for comparison. The product properties can be compared according to its type. For our purpose model we will consider two distinct set of products. First type of product which we have considered is electronic products. The electronic products are those which depend heavily on the gates and switches. Second category of product which we have considered will involve electrical products. These are the products which depend heavily on the voltage and current. There are certain characteristics which are optional and which are required in both the electrical and electronic products.

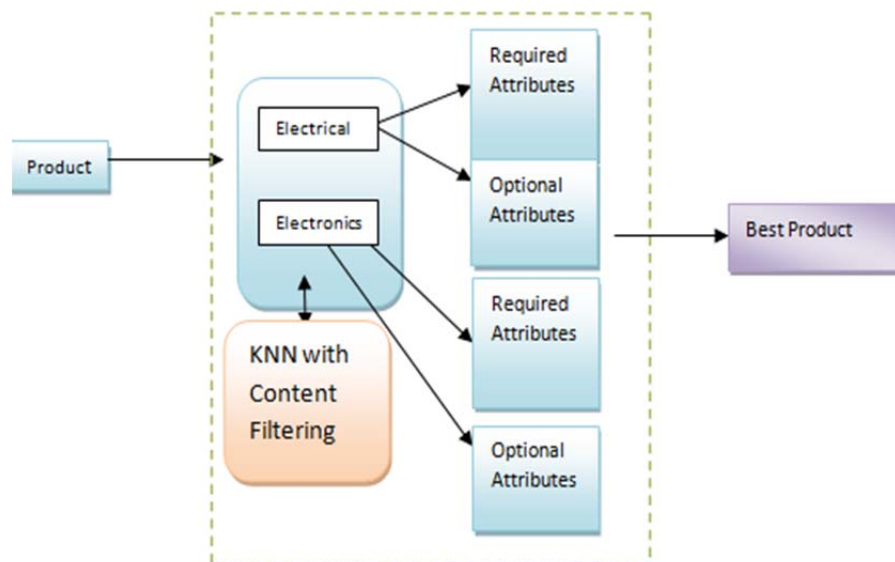


Fig 1 showing the products to be promoted

Some prerequisite for the electrical product under various categories to be Green Will be listed as follows

A. Required Attributes for Electrical Equipments

1. Compliance with European RHS(Restricted and Hazardous substances) Directive
2. Amount of Mercury
3. EU Battery Directive
4. Use of plastic casing
5. Declaration of product weight
6. Use of single recyclable plastic type per plastic part
7. ENERGY STAR®
8. Marking of plastic components
9. Provision of product take-back service

B. Optional Attributes for Electrical Equipments

1. Use of non mercury plastic contents
2. Amount of chemical utilized within the product
3. Use of bio base content material
4. Ease of disassembly of the product

C. Required Attributes for Electronics Equipments

1. Compliance with provisions of European RHS Directive upon its effective date
2. Reporting on amount of mercury used in light sources (mg)
3. Elimination of intentionally added SCCP flame retardants and plasticizers in certain applications
4. Declaration of postconsumer recycled plastic content (%)
5. Declaration of renewable/bio-based plastic materials content (%)
6. Identification of materials with special handling needs
7. Elimination of paints or coatings that are not compatible with recycling or reuse
8. Marking of plastic components
9. Availability of additional three year warranty or service agreement
10. ENERGY STAR®
11. Provision of product take-back service

D. Optional Attributes for Electronics Equipment

1. Elimination of intentionally added cadmium
2. Low threshold for amount of mercury used in light sources
3. Elimination of intentionally added mercury used in light sources
4. Elimination of intentionally added lead in certain applications
5. Elimination of intentionally added hexavalent chromium
6. Minimum content of postconsumer recycled plastic
7. Higher content of postconsumer recycled plastic
8. Reduced number of plastic material types

The above said attributes has to be satisfied in order for the product to be Green. Optional parameters may be skipped but required parameters have to be satisfied.

IV. CONTENT BASED SYSTEM

Content Based System consists of rules that can be used in order to determine whether the product is Green or not. The recommender systems will go to promote the products which satisfy the conditions specified. The data set will be prepared with in the fuzzy system. The Dataset will be compared against the attributes of the new product. The conditions if satisfied than the product are rated as Green. The membership functions will be defined in order to determine whether the given product lie in a given range or not. The rules regarding the fuzzy system (Patil & Dhamakale, 2011) will be as described as follows

- a. Compare the input variables with the membership functions to determine the values of the Label. This is also known as fuzzification.
- b. Combine the membership values with the premises values to get the firing strength.
- c. Determine the result of the conditions specified.
- d. Aggregate the results so that desired output in terms of Green Products can be generated.

The recommender system will use the above rules of fuzzy logic in order to determine whether the product qualified to be Green or not.

V. PROPOSED SYSTEM

The products which are consider for recommendation will according to the satisfying constraint. The constraints are listed in terms of the criteria. Every constraint would compromises of number of conditions. Every condition within the criteria must be satisfied in order for the complete criteria to be satisfied. The model will be listed as follows

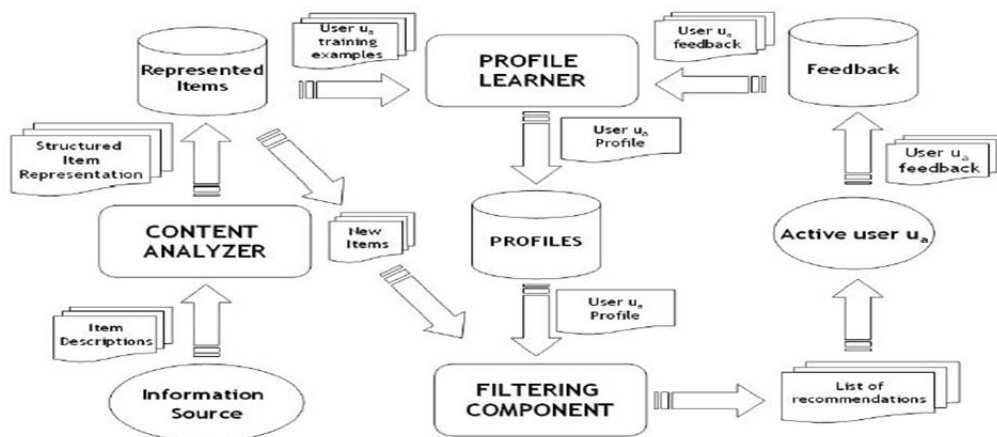


Fig 2 showing the proposed system

The components of the proposed system will be as follows:

a. Structured Item representation

The structured item representation is used in many software development projects. The structured data representation will help in expressing the information about the product which is easy to understand without going into the detail. In other words abstraction is provided by the use of structured representation.

b. Content Analyzer

The content analyzer will help in comparing the feature of the product with the attributes of the desired product. The match result could be positive or negative. The positive result items will be accepted and negative result items will be rejected.

c. Profile Learner

The profile learner will help in gathering the attributes of the new green products if not already present within the data base. The new attributes then can be used to determine the other green products if necessary. Profile will be stored with the help of profile learner.

d. Filtering Component

The filtering component will use the concept of content filtering in order to determine the Green Products and also matching will take place in order to determine the list of similar products.

e. Profiles

The profiles will indicate the characteristics of the products. These characteristics of the products will be recorded so that they can be compared later on if required.

f. Feedback

The proposed system suggests two ways communication. In order to achieve that user feedback is compulsory. User feedback is stored within the database. This feedback will be used to determine the performance of the current product. The feedback can be positive or negative which will be used to modify the exiting level of greenness of the product.

The proposed system will mainly use two functions which will provide the information about the Green Products and how satisfied the users are with the product. One function of the proposed system is to make designers and specialist understand the consumer preferences and how satisfied the customers are with the Green products. Other function of the proposed system is to provide the Green products information retrieval system. The level of greenness will be displayed on the internet. The user opinion will also be considered in this case. The user opinion will be based upon the experience of the peoples who have used these products. Accordingly level of greenness of the product will be varied. We propose a recommender system based on the content filtering which will consider the ingredient of the products in order to determine level of greenness. The product with the higher degree of greenness will have higher probability of being promoted. We shall collect the information about the Green Indices and Green Score of product to be promoted. The collaborative filtering can also be used which will record the reviews suggested by the users. The users who have used the product will respond according to their experience of using the products.

The algorithm for the proposed approach is as under

Algorithm C-KNN

- Take the input from the Dataset
Input_i=Dataset
Parameter="Likes"//Prediction on the basis of Likes from dataset
 - Apply filtering mechanism to obtain filtered values in distinct classes
Class_i=Filter_{Condition}(Input_i)
 - Apply KNN to group nearest neighbours together
In k-NN grouping, the yield is class participation. A question is grouped by a larger part vote of its neighbors, with the protest being doled out to the class most normal among its k closest neighbors (k is a positive number, regularly title). On the off chance that k = 1, then the protest is essentially allocated to the class of that solitary closest neighbour.
In k-NN relapse, the yield is the property estimation for the question. This esteem is the normal of the estimations of its k closest neighbors.
 - Predict Group with most "Likes"
-

Before establishing our Green Recommendation System we first collect the information about the degree of greenness of the product. We define N green indices which will have degree of greenness $\{g_1, \dots, g_p, \dots, g_j\}$. Here each degree is associated with the set of criteria's R_j . the degree of Greenness will be influenced by these factors. The products P_i which satisfy highest degree of criteria's will have highest degree of greenness. We will also maintain a qualified vector $C_{ij} = \{C_{ij}^1, C_{ij}^2, C_{ij}^3, \dots, C_{ij}^R\}$ where $C_{ij}^r=1$ is the particular criteria for the Product P_i is satisfied. Otherwise $C_{ij}^r=0$ if no criteria is satisfied. In case some criteria's are satisfied then Green indices will be calculated as

$$W_{ij} = \sum_{i=1}^{R_j} C_{ij}^r / R_j$$

The recommendation system will take into consideration vector space model. The weight matrix will be constructed for this purpose. The matrix will display the degree of the greenness of the product. Higher the degree more probability will be that product is promoted by the recommender system. The matrix will be shown as follows

Table 1 showing the product with green indexes

	g_1	-----	g_j	-----	g_n
P_1	W_{11}		W_{1j}		W_{1n}

P_j	W_{j1}		W_{jj}		W_{jn}

P_n	W_{n1}		W_{nj}		W_{nn}

User can decide the Green product to be chosen by looking at the matrix. The matrix describes the degree of Greenness. User can select the product according to his or her own preferences. The preference vector (V_j) is also going to be maintained. The user will select product based on g_j and set $V_j=1$ else $V_j=0$ which will indicate that user is not interested in the index g_j .

In order to determine the product according to the user preference weight threshold (W_j) will be maintained. User rating can be assumed to be from -10 to 10. Accordingly weight threshold will be calculated as

$$W_j = V_j / 10 * W_j^{max} \text{-----Eq-1}$$

W_j^{max} is the maximum weight corresponding to the row matrix. Highest the preference degree higher will be the threshold.

Five rating levels are provided if user rating is very poor it is represented with -10, if user rating is poor then it will be represented with -5. No opinion will be represented with 0. If rating is good then it will be represented with 5 and very good rating will be represented with 10.

There are following rules which must be followed in order to check the rating and also to improve the rating depending upon the reviews of the users who are using these products

Rule 1: The new rating observed is better than the existing rating. If this is the case then existing rating will be replaced with the new ratings. If the existing rating is negative and the new rating is positive then the new rating is double the existing rating.

Rule 2: The new rating is worst and hence the existing rating is reduced. The rating will be doubled if the new rating is positive and the existing rating is negative.

Rule 3: The new rating is equal to the existing rating then the existing ratings are retained.

Locality Sensitive Hashing procedure including K nearest Neighbor is listed as under

This algorithm will produce similar products with similar hash code

- For each product P_i
 - a) Generate a K bit Hash Code
 - b) Insert Product into Hash Table
 - c) Check for Collision(Element Exist) check only within the particular bucket
- Check for near duplicates
 - a) Similar hash code \neq same bucket
 - b) Repeat the steps for all the hash tables presented

Both the above algorithms are used in order to determine the green products and listening all the matching products also.

VI. Conclusion and Future Work

The proposed system is a recommender system which will go to promote Green products. Various parameters are considered in this case for input. The parameters which we have considered are Eco, Organic, Stars, Power and Recycled. The output which is generated is divided into parts- Green Products and Not Recommended. If the product is identified as Green then it will be promoted using our recommender system. Otherwise the recommender system will reject the product.

Although the performance of the recommender system is good and it will detect the Green Products which can be promoted using Recommender system but still there could be more accuracy which can be used in order to detect the Green products.

VII. References

- [1] Baltrunas, L. (2011). Context-Aware Collaborative Filtering Recommender Systems, 4(April), 172. <http://doi.org/10.1561/1100000009>
- [2] Barreda, A. A., Bilgihan, A., Nusair, K., & Okumus, F. (2015). Generating brand awareness in Online Social Networks. *Computers in Human Behavior*, 50, 600–609. <http://doi.org/10.1016/j.chb.2015.03.023>
- [3] Berka, T., & Plöbng, M. (2004). Designing Recommender Systems for Tourism. *Proceedings of ENTER 2004*.
- [4] Hoic-Bozic, N., Holenko Dlab, M., & Mornar, V. (2015). Recommender System and Web 2.0 Tools to Enhance a Blended Learning Model. *IEEE Transactions on Education*, 1–1. <http://doi.org/10.1109/TE.2015.2427116>
- [5] Lee, Y.-L., & Huang, F.-H. (2011). Recommender system architecture for adaptive green marketing. *Expert Systems with Applications*, 38(8), 9696–9703. <http://doi.org/10.1016/j.eswa.2011.01.164>
- [6] Patil, S. B., & Dhamakale, S. D. (2011). Fuzzy Logic Approach with Microcontroller for Climate Controlling in Green House. *International Journal on Emerging Technologies*, 2(1), 17–19.
- [7] Resnick, P., & Varian, H. (1997). Recommender systems. *Communications of the ACM*, 1–21. <http://doi.org/10.1162/153244302760200641>
- [8] Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., & Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. *Procedia Computer Science*, 1(2), 2811–2819. <http://doi.org/10.1016/j.procs.2010.08.006>
- [9] Zheng, X., Member, S., Chen, C., Hung, J., He, W., Hong, F., & Lin, Z. (2015). A Hybrid Trust-based Recommender System for Online Communities of Practice. *IEEE Transactions on Learning Technologies*, 1382(c), 1–13. <http://doi.org/10.1109/TLT.2015.2419262>