Content Based Apparel Recommendation System for Fashion Industry

Illa Pavan Kumar, Swathi Sambangi

Abstract Present apparel e-commerce system that encourage online shopping has major issues to deal with catalog based online shopping. As there is a lack of customized services, the users may face difficulties to find discrimination over different types of retailers available on electronic product catalogs, they may also be confused with complex navigations that redirect to other pages based on their selection. This drawback can be overwhelmed by following suggestions on categories that they have chosen or from the products that they have already viewed. Multiple number of online marketing companies around world-wide, has been practicing the naive method for apparel marketing website. This paper aims to simulate this recommendation system on real world data set taken from the marketing giant, Amazon’s Product Advertising API, in a policy compliant manner by following the procedure in three steps: Analyzing the data to select the pivot for the recommendation system, Data preprocessing to remove invalid sections and to implement and find appropriate choices among the techniques like Bag of Words (BoW) and TF-IDF for better recommendations.

Index Terms: Amazon’s Product Advertising, Bag of Words, TF-IDF.

I. INTRODUCTION

The industry for apparels has a major role in the development and globalization of countries as it has been incorporated into the world economy. Fashion industry involves huge supply demand chain of designing garments ,and production of sales. In this contemporary world, there has been a phenomenal change in supply demand chain. Since 2008, there is a consistent increase in the sales of garments by $3.3 billion each year, and the sales of global garments has achieved $1.25 trillion in 2012, that is an sign that there is a great and a fast growth in development of industry for apparel.

The recommendation technology for apparel industry, as an emerging technology, has attracted wide attention of scholars. It is a well known fact that the apparel recommendation depends on operations which were performed manually. To help the customers in selecting the apparels with their own customization, the sales person could do with to a good recommendations to the customer so as to increase the satisfaction rate in customers about the products, and make the customer to be enthusiastic in purchasing the products, to understand the sentimental analysis of the customers during shopping and finally to increase the product of sales.

On the other hand, a sale person can face a lot of complexity in order to identify with thoughts of customers’ and after that to recommend with an apparel that reaches the customized needs of a customer. As a result, there is a necessity and a great purpose to stumble on a set of objective indicators, rather to concentrate on subjective opinions, for evaluating role of recommendation technology in apparel industry.

Thus recommendation system plays a vital role in e-commerce applications like movies, shopping, tourism, TV, taxi.

1.1 What is a Recommendation System and Recommendation Engine

One who wants to be a good merchant, need to understand the concerned priorities of customers. If a Recommendation system can be deployed into a ecommerce application especially like apparel industry, the technology could guide the customers with high quality of recommendations and give contentful results for the customer to buy and increase the sales with good profits.

A Recommendation System can be defined as a technology that gathers data about personal interests and taste of a customer and give suggestions on product sales as per their preferences. The so called recommendation system collects the information from the user by following two ways: either implicitly or explicitly. Implicit method of collecting information involves close observation or monitoring of customer's purchased products where as explicit acquisition of customer preferences involves observation on previous ratings or history. Collecting the information from the customer needs could be possible by using a data filtering tool which use intelligent algorithms and customer's preferences data as input, in order to recommend the most interesting products to the concerned customer. We can consider the recommendation engine as a automated sales person who displays a product with similar products that a customer is interested in. The Recommendation Engine consists of algorithms which are well trained to increase the product sales as per consumer needs.

1.2 Definition Recommendation engine and its examples:

The Research paper entitled with the name “Recommendation Systems: Principles, methods, and evaluation”, F.O.Isinkaye defined Recommendation Engines/Recommender Systems as follows:

“Recommendation Engines/Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of a large amount of dynamically generated information according to user’s preferences, interest, or observed behavior about the item.”

Revised Manuscript Received on August 05, 2019.

Illa Pavan Kumar, Department of Information Technology, VNRVJET, Hyderabad, India. (illa.pavankumar@gmail.com)

Swathi Sambangi, Department of Information Technology, VNRVJET, Hyderabad, India. (ssambangi555@gmail.com)
Recommendation Engines/Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user’s profile” Recommendation Engine—Examples A few examples of that make use of recommendation are observed in the following Facebook—“PeopleYouMayKnow” Netflix—“OtherMoviesYouMayEnjoy” LinkedIn—“JobsYouMayBeInterestedIn” Amazon—“Customerwhoboughtthisitemalsobought…” Google—“VisuallySimilarImages” YouTube—“RecommendedVideos”

II. RELATED WORK

As there is an enormous information can be accessible from the internet and a tremendous rise in the electronic content, the people may possibly face a challenge to access the overloaded digital information and retrieve the useful and relevant data from the ocean of information available on the Internet. The problem of tracing and accessing the relevant and useful data has been partially solved by information retrieval systems but there is a lack of users preferences and priorities of the data. Thus there is a demand for a system that filters the information and can solve the problem of accessing personalized content from overloaded information available on the internet. Online sales applications make use of Recommender Systems to display the products or items based on customer interests. This has lead to the discovery of Recommender Systems with an ability to predict and suggest particular product or item based on customer preferences and priorities.

2.1. Evolution of Apparel Recommendation System:

There is a magnificent development in the Apparel consumer market since the past few decades with a progressive challenging issues in the market. In order to resolve the challenging issues in Apparel retail industry has concentrated on apparel recommendation system. Apparel consumer market with the help of the intelligent recommendation system could have more benefits economically.

2.1.1 History of Apparel Consumer Market:

Prior to the discussion of how e commerce applications for apparel industry g has came into existence with apparel recommendation systems, there is a need to understand the history of Apparel Consumer Market. A brief history of Apparel Consumer Market has a great relation with the history of Britain in the late 17th century. The role of Urbanization and industrialization in the great Britain has introduced the customers to buy the clothes as per their preferences. Thus the demand for Apparel Consumer Market has began to rise as the customer group were interested to buy readymade clothes. Likewise the Apparel Consumer Market has changed supply demand chain from raw cotton exports to readymade clothes. Eventually the raise in fashion technology and globalization of the apparel consumer market, the products were made available for the customer through e commerce applications, so that customers can buy the clothes without visiting apparel retail store. This has lead to store lot of digital information on the internet which throws some challenges for the customer to pick the clothes of their preferences within a short time. To overcome those challenges and help the customer to show the products of their interest Apparel Recommendation Technology has come into existence to filter the digital data by using some intelligent algorithms as per the customer preferences to increase the product sales and to attain customer satisfaction.

2.1.2 History of Apparel Recommendation Technology:

The history of recommendation technology has been introduced fairly with the history of computing itself. The first automatic recommender Technology has been introduced as a computer based librarian, to generate recommendations on book preferences given by the user, by grouping users into “stereotypes” depending on the results obtained after conducting a short interview with the user. Later in 1990s, in order to handle huge amount of information available online, collaborative filtering, a type of recommendation system has been introduced to deal with the data. The first collaborative filtering recommendation system called Tapestry, a manual collaborative filtering system, allowed users to select the items needed by giving recommendations as a result of queries that a user quote for selection of items from a large domain of information available. Later another type of recommendation system namely Automated collaborative filtering system has been introduced to give recommendations as a result of relevant reactions given by the users. This technique has been introduced by Group Lens, to understand the interest a specific user about the Usenet articles.

The past decades has renewed the importance and achievements of Recommendation system in apparel consumer market. To promote a customer to select an apparel without wearing it, an apparatus has been introduced by Lam in 2003, for personalized apparel selection by the customer. As the apparatus did not give successful results as expected, a color planning method has been proposed by kobayashi in 2004, for coordinating with apparel colors. This method helped the customers to select the apparel depending on their color of their preferences for their outfit. Later in 2009, an expert system namely fashion mix- and-match method has been proposed by Wong etal. This expert system automatically recommends the customer and facilitate their preferences for a mix and match apparel fashion. This method mainly focus on collecting the data and evaluating the decisions from the fashion designers, and help the customers in apparel coordination. This expert system could be able to store the captured data in the form of information on the web. Thus recent developments in apparel recommendation technology has shown that the system has been evolved with a variety of intelligent algorithm used and apparel recommendation technology has turned as intelligent apparel recommendation technology.

2.2. Design of Apparel Recommendation System:

There are several techniques are used for apparel recommendation technology. Considering the satisfaction rate among the customers about the products, not only that but also the customized preferences of apparel by the customers, from the browsing history of the clothes by the customer are some of the major issues to be accounted for recommendation technology. Additionally, a few researchers has noticed that the statistics of clothes taken from the browsing history shows that,
the recommendation on apparel can also be done with the accessories and weather conditions at that particular time.

To achieve various requirements, there is a need for an intelligent recommendation system, for apparel industry to suggest an apparel that reaches customers preferences depending upon various ideologies that were studied during from the wide research. Iwata et al (2011)[18], presented an recommendation system, by making use of full body photographs from fashion related magazines. Similarly, Sha et al(2016) [19] have analyzed how to recommend the most suitable apparel design scheme for a particular customer, by extracting various several features from the images to investigate or examine the contents of the apparel like type of sleeve, type of collar, pattern of the fabric and any other various attributes.

To design an apparel recommendation scheme for a particular customer to choose an apparel of their own preferences, the themes and shapes related to fashion technology and the perceptions of professional fashion designers is incorporated. The theory of recommendation technology has been introduced in the middle 1990s. The wide research that has been done in this recommendation technology shows that various varieties of algorithms were developed from time to time. To create a profile for apparel recommendation, CRESA can be considered as an example that joins attributes related to text, features that gives visual information, and finally the customers attention towards the model of the apparel displayed in the catalog.

A new collaborative filtering strategy which uses the attention of the customers' view that distinguish the images and also mitigate concentration on new item, has been suggested by Nogueira et al (2015)[20]. For predicting the best recommendation ranks about the products displayed, understanding the customers views about the products, Scholz et al(2015) [21]suggested exponential utility function for the recommendation technology.

Supplementary clothing intelligent tools and virtual fitting measurements has been built up as quickly as it can with the fast growth of Internet technology. 3D graphics technology was applied first to generate and simulate the virtual store, by Cordier et al(2003)[22]. The interactive 3D virtual fitting room system, has been proposed by Li et al.(2011)[23] that helps to change customers preferences as per the hairstyle and accessories used by the customer and suggest the customers for selecting proper clothes according to their preferences. On the other hand, the virtual fitting research products are continuously acts as cutting edge for developing recommendation technology. In reality, the main purpose systems in apparel recommendation technology is just to exhibit apparels, and only a sample of the fitting effect can be viewed by the customer. The search for the products is a tedious and cumbersome task if there is no effective recommendation technology is associated with online apparel industry.

An interactive clothes fitting system has been presented by Zhang et al.(2008)[24], which is used to identify the human eyes vision with respect to similarity index in clothes through front view outfit images. The Closet Application was suggested by Limaksornkul et al. (2014)[25] for documenting the type of accessories and statistics of clothing which are utilized very often to customers as per the statistics that has shown form their history of product purchase. There is a great significance of recommendation technology as it permits the customers to find a suitable apparel according to their preferences as fast as it can and also it helps to increase the sales of apparel industry. However the proposed methods are mainly based on the subjective views which that pay no attention to the objective data. With an intention to deal with this problem, we propose an evaluation method for apparel fashion industry to recommend the products.

2.2.1. Structure of Apparel Recommendation System

The past data that is collected from the intelligent algorithm implemented by the recommendation technology, can be utilized to know about the facts that which products are being liked by the customers and were brought, and which were preferred to be added to wish list from which sentimental analysis of similar consumers can be made. Previously the entrepreneurs used to collect the data from the customers with the help of membership cards and feedback questions. This type of traditional method cannot assure the data authenticity as there is a consistent up gradation of the data according the alterations made by psychological aspects of consumer. Presently, the personnel who deals with e-commerce applications for their business purpose concentrates on recommendations from virtual member registrations, consumption tracking, information gathered from browsing history. Even though there is lot of effort has been made to know all the information about customers preferences and the interests, the results obtained were not considered as ideal as some consumers cannot give the complete or genuine information with a fear of leakage of their private information on the internet which can leads to face some security vulnerabilities. So there is a need for proper structure of the apparel recommendation system. In this paper we followed the illustrated structure given below.

The structure contains a customer interface and machine inference tools, knowledge base, storage of data. The interaction between customers and system is being made with human interface with machine. The information obtained from this module is used to invoke the knowledge about customer preferences and use it as a knowledge base for machine to apply inference rules. Finally the facts and rules are obtained from the knowledge base and the data is stored in the database for further recommendations to be made.

2.2.2. Taxonomy of Recommendation System:

Recommendation System have several techniques to give suggestions for the customers to have the results of apparels according to their preferences. Sometimes recommendation system can also be referred as utility function that anticipates the utility rating of an apparel for specific customer. According to the growing literature in the research of recommendation systems, there has been several types of Recommendation Systems are proposed.

a. Content-based recommendation system:

The content based recommendation system works according to the data that has been provided by the customer either explicitly or implicitly. Once the data is collected then the recommendation system raises the suggestions to the customers to select the similar items whose characteristics are likely to be similar with the past preferences of the customer. In turn the content - based recommendation system can be an approach that analyses description of the content only or it is just building user profile and item profile from user rated content to give suggestions for the customers.
b. Collaborative recommendation system:
This type of recommendation system provides recommendations for the customers according to the polls or reviews of the customer community. This can have two approaches. In the first approach the main idea is to encourage the customer to view the new products on the similarity of other customers tastes with the current customer. And this is named as User based collaborative recommendation system. The second approach of this type ,targeted to give the recommendations according to the item similarities of the target user community. This approach is named as item based collaborative recommendation system.

c. Demographic recommendation system:
This approach makes use of demographic profile of the user to provide recommendations about the items that could reach customer preferences. Demographic profile of customer consists of the data like customer's age, gender, languages known, education, employment status, ownership of home and location.

d. Utility-based or knowledge-based recommendation approach:
The recommendations to customers can be given in this approach by considering the knowledge about the items whether the characteristics of the item could reach the specifications or the preferences given by the customer or not. This type of recommendation system can be drawn into two categories namely case-based and the other is constraint-based. In case-based approach, Information about the similarity metrics of the user and items are used to give suggestions. In constraint based recommendation system deals with a set of rules that guide to relate user requirements with item characteristics.

e. Community-based recommendation approach:
The fast growth of social networking derives the recommendation system to make use of social networking analysis of data to gain the ability of recommendations on the system. As the trust factor is more important in this approach, the customer pays more attention on the friends reviews or recommendations on social networking about a particular item rather than concentrating on anonymous comments about the item. This approach is purely based upon affirmative opinions in social networking, online friends communities and chatting in group in which an item can be recommended depending on its social tags and social bookmarks obtained during social interactions of the customers.

f. Context aware recommendation system:
The mentioned recommendation system approaches does not recommend the items or results depending upon the customers current preferences as they are more interested in considering the long term preferences of the customers to recommend the products. This approach makes a difference among all other recommendation techniques as it considers the customers short time preferences by examining the current context like weather, time, region, behavior and emotional state etc.

g. Hybrid approach
This approach generates recommendations depending upon the approach generated by the combination of above mentioned recommendation techniques. By using this approach, one can overcome the disadvantages occurred in any one of above mention recommendation technique by combining that with other approaches. Hybrid recommendation system gives the results are more precise and have high performance index.

III. SYSTEM REQUIREMENTS
For coding purpose we have chosen Python Programming language. In accordance with the most recent TIOBE Programming Community Index[26], Python is one of the top 5 popular programming languages as of June 2019. With Python one can run the Programs directly without compiling it into binary language commands and it is easy to represent real world entities using Python because of its plain and easy-to-read syntax. Python dominates other programming languages with its rich set of libraries that deal with Big Data Processing, Scripting & Automation, Data Science, Artificial Intelligence, Software testing, Computer graphics and web development. It has numerous ready-made packages to carry-out scientific computation based on the requirement in above mentioned areas. For this paper, as per the requirement we have imported Numpy, Pandas, NLTK Scikit-learn, Matplotlib packages from Python libraries. Glimpse of the above packages is given below.

NumPy: This package in python is designed for efficient scientific computation[27]. NumPy enhances Python with dominant data structures, execute multi-dimensional arrays and matrices. These data structures assure efficient computing with matrices and arrays. Numpy usage even aimed at large array processing.

Pandas: Pandas stands for “Python Data Analysis Library”[28]. It is written for Data handling and analysis. Pandas is quite a game changer when it comes to analyzing data with Python and it is one of the most preferred and widely used tools for data analysis. Pandas DataFrame method Data manipulation and analysis easy when compared with other methods.

NLTK: It is Python Natural Language Toolkit to deal with human language data[29]. It is helpful in Natural language Processing. It is intended to split the natural language in to set of tokens, to identify stop words like “the”, “is” “an” and to remove them from given input text. NLTK contains in built list of possible stop words from English language. NLTK tokenize and corpus modules are used for splitting and identifying stop words from naural language.

Scikit-learn: It is a Python package used to solve machine learning problems though a steady interface[30]. It offers an extensive collection of algorithms for classification, Regression, Clustering and dimensionality reduction. It also offers modules for model selection and Preprocessing(Feature extraction and normalization).

Matplotlib: It is a plotting library used to generate 2D graphics with Python. The generated plots features are appropriate to put the results in a research paper[31]. Merely it is a Data visualization tool to visualize data through Bar graphs, Histograms, pie plots, scatter plots etc.. Data visualization can be accomplished by importing pyplot module form matplotlib library. Otherthan the above mentioned python libraries, for interactive computation purpose we made use of Jupyter Notebook: An open source software application based on web browser[32]. It supports numerous Programming languages like R, Python, Scala etc.
IV. METHODOLOGY

To build a recommendation system, we will take the information from various internet sources. In this paper for apparel recommendation system, we have collected the data from Amazon website. The information that is extracted from Amazon sometimes may be unstructured data, which after processing could not be able to produce relevant results that can be used by recommendation engine. In order to compare the feature extraction techniques, we have tried the following phases for the information that is gathered as a prerequisite procedure.

4.1. Loading the data:

As the information gathered is considered as a huge dataset, it is difficult for any learning algorithm to work and produce most accurate results. Hence there is a need of preprocessing techniques to be applied on the data. Before transforming the information as a flexible input for the learning algorithm, there is a need to gather information properly. The data used for the recommendation engine for apparel recommendation system has been obtained from the Amazon API in a policy compliant manner. The data extracted from Amazon Product Advertising API using REST API and the data is extracted in JSON format. The data is loaded using pandas library in python and stored in a dataframe which can be sent for further processing and also the dimensions of the data is noted.

4.2. Preprocessing of the data extracted:

After the data is imported into dataframe, then it undergoes for preprocessing phase. In this step, data frame is sent for preprocessing to eliminate the redundancies and irregularities in the data. Preprocessing of data is mainly done in three steps:

a) Tokenization:

The procedure of dividing the given input data into part of individual pieces is called Tokenization. In any data frame, punctuation marks, words and numerical values and others can be considered as tokens. In this paper tokenization is performed by using Natural Language Toolkit available in python library. For the apparel recommendation system, tokenization is performed for removing data of products with missing price and color, removing products with few words in title by using near title duplication.

b) Stop word removal:

In English, there are the majority of the words and some punctuation symbols does not play a significant role and does needed to a great extent for the algorithm in producing the results. Such words can be termed as stop words. For example words like in, nor, but, am, at and so on and punctuation symbols like ' ', '', #, $, @, %, ^, & and so on. In this regard, it is required to eliminate stop words for minimizing the process time of the algorithm. It is achievable to remove stop words in python by using the following steps:

i. The list of stop words are downloaded from nltk which is a collection of libraries that are necessary for the statistical processing of any natural language. We can also define additional stop words if any, split the additional stop words and we can add the specified words to nltk stop word list.

ii. Now the stop words list is created as tokenized words and are passed as parameters for processing including the special characters.

iii. In this step all the letters are converted to lower case.

iv. The words which are identified as stop words are removed and replaced with blank spaces.

c) Stemming/Lemmatization:

This step is used to find prefixes and suffixes for a root word. If a customer want to search for an apparel, its root is taken as parameter to search and all the types that start with that particular word are displayed for the customers which improves the performance of the recommendation system. The process of stemming is mainly done to map all the set of similar objects together.

4.3. Feature Extraction:

Most of the data which is gathered from web sources is unstructured in nature. The huge amount of data which is an important source to analyze and reuse for further research applications can be processed by humans up to some extent even with difficulty. But it is difficult for us write a learning algorithm to process and analyze the data that we have gathered. In this context feature extraction comes into play as a prominent role as it extracts the important features that are necessarily participate in analyzing the text in classification of the documents by identifying related features.

There are several feature extraction methods available for the procedure of bringing a list of words together to enhance the analysis of the information gathered. This emphasizes the working nature of recommendation engine to provide the better recommendation of items for the customers. This paper highlights the performance metrics of the three feature extraction methods which are applied on Amazon apparel data to design a good recommendation system for apparel application. We have worked with BagOfWords(BoW), TF-IDF and IDF. The later section in this paper will explain the three feature extraction methods and compare the results obtained to analyze the performance of the three methods.

4.3.1 BagOfWords(BoW):

The bag of words is the straightforward technique used for feature extraction method. In this method, an instance of sentence is taken and it checks for the for the presence feature set that is predefined. It does not check for the frequency of the occurrence of the word but it just check for the presence of the word in the given instance.

In this work, bag of words is implemented on text that is used for searching an apparel. For example, the customer is using the following keywords for an apparel to select, then the following table illustrates the working nature of BagOfWords(BoW).

Main Title of the product: Unibelle blouses casual t-shirts long sleeve cotton topswine reds. For every word that is present in the main title of the product compared with the words in the searching title, then the similarity index for the word is given as ‘1 else 0’. Search Titles samples:

i. Unibelle blouses casual t-shirts long sleeve cotton tops wine reds

ii. Unibelle blouses long sleeve cotton shirt tops women black

iii. Rock cotton long Sleeve Tunic

iv. Long sleeve mock neck top

v. Unibelle crewneck long sleeve cotton shirts tops women wine red

vi. Women Oxford shirt Long Sleeve
4.3.2 TF-IDF:

This is a feature extraction method in which the frequency occurrence of a domain specific text is calculated to judge the importance of the word in the information gathered. The transformation of the words can be easily done if we could recognize the signature of words. TF-IDF is a combination of two feature extraction methods: Term Frequency and Inverse document frequency. In this type, term frequency calculates the importance of the word in the text where as Inverse document frequency concentrates to find the signature of the word in relevance with other documents or searches.

After applying this extraction method on Amazon data, we got the tf-idf values (mentioned in the result section) for the given searching string when compared with the main title.

V. RESULTS:

5.1. Bag of Words:

5.2. TF-IDF:

5.3. IDF:

VI. PERFORMANCE ANALYSIS

The experimental analysis of feature extraction methods proves that the combination of TF-IDF gives best results. And following are figures shows the shortest Euclidian distances.
VII. CONCLUSION

The primary goal of this paper is to recommend a one to one text based apparel recommendation engine, with data that was obtained from the Amazon API in a policy compliant manner. The Amazon recommendation systems acts as an efficient technology that triumph over huge information available on internet. This technology helps the customer to choose the products of his/her customized products and which increases the growth rate of customer satisfaction. In this regard, this paper also concentrated the feature extraction techniques and compares them with experimental results to suggest the best method for better results. Thus the one to one apparel recommendation system investigate the best methods to retrieve the information and features that helps to understand the online retailer about the sentimental analysis of the customers to increase the production of sales and customer satisfaction.

REFERENCES

27. https://www.ioibe.com/ioibe-index/
30. https://maplotlib.org/
31. https://jupyter.org/
AUTHORS PROFILE

Illa Pavan Kumar completed B.Tech(CSE) from Kakatiya University and M.Tech(CSE) from K L University. His research interests includes Cloud Computing and Machine Learning. He has 6 years of teaching & research experience.

Swathi Sambangi completed B.Tech(IT) from JNTU, Kakinada and M.Tech(CSE) from JNTU, Kakinada. Her research interests include Cloud Computing and Machine Learning. She has 7 years of teaching experience.