Product Recommender Chat Bot

Neera Sanjay Agashe
Computer science and Engineering,
Government College of Engineering,
Aurangabad, India

Abstract—**There has been number of researches in the field of data analytics to recommend the products, and many of them are successfully recommend the desired content what user is asking for. There are several domains such as e-commerce, movies, musical instruments, websites, books, etc. where recommendation system has its own significance. The outcomes provided by these domains has the scope of improvements so that final result will reach to user's satisfaction. This research use the recommendation system and provide the output desired by user. Research is introducing the chat bot which recommends the product to the customers as per their requirement. The chat bot is basically order taking with minimal user input and suggested the appropriate product. This can be done in large scale, but here I am using niche perfume database as product. Customer provide the details about the perfume through the chat bot. And according to user description it will recommend the related products. Doc2Vec, Latent Semantic Analysis, and Sentiment Analysis come together to make relevant recommendations in a chat bot interface. Further, the cosine similarity is used to recommend the perfume. For this, cosine similarity of search query is matched with the average of the cosine similarity of LSA and Doc2Vec document embeddings.**

Keywords—Doc2Vec, Recommendation System, LSA(Latent Semantic Analysis)

I. INTRODUCTION

Internet shopping is productively growing in today's world of e-commerce. So, Product recommendation systems also has a chance to be developed. Since, users need a relationship between them and system. When relationship get build up then user get personalized care and attraction. System just not observe and analyze shopper behavior but also attracts them to come and buy again. Recommendation system reduces the boring task of users to search in an endless category what they want. Instead, they use the conversation as a filtering system, bringing the product to the customer. Online shopping has many advantages but their limitations and disadvantages also must be considered. User buying the product and what user is asking for doesn't match every time, this may lead to disappointment. As the needs of the users keeps changing day by day, the improvement of existing functionality of these systems has become a crucial factor. By analysing history of internet shopping, there will be huge demand for recommendation systems in near future.

Customers buy according their choices, moods, events, etc. And the perfume is one thing that everyone loves it. Also its can be gifted. So, this research will recommend the perfumes according to customer's moods, likings, etc. Customer just has to write description of perfume which he/she wants to buy. For example – I want something to wear at beach or pool. This description is enough to recommend you the appropriate perfumes you are exactly looking for.

II. RELATED WORK

The huge amount of digital data is exponentially growing day by day leading to the difficulties for users to search the required data. Search engines like Google solve the problem of data availability but the issue of personalizing content's to the user remains as it is. Hence, necessity of recommendation system is increased to overcome these problems. These system worked by filtering the huge amount of data and provide the information according to user's requirements. For performance evaluation ResQue Recommender Systems Quality of Users Experience framework is used. This evaluation was based on measures that include qualities like the usability of the system, interaction quality, user's satisfaction and user's behaviour etc., which help to know how user is reacting to the system.

E-commerce platforms are framed using recommendation system. This system try to recognise customer's behaviour and then recommend the products according to their interest (Schafer et al.; 1999a). The wildly growing shopping websites such as Amazon, Levis has their own way of recommending products. For example, Amazon5 recommends user items that have different features like Customer who viewed this item also viewed, Customer who bought this item also bought. It suggests the Customer what else he can buy with already bought products.

Mooney and Roy (2000) developed a system named as Content based book recommending using learning for text categorization. It concludes that Content based methods and machine learning algorithms can produce accurate recommendations. A prototype, Learning Intelligent Book Recommending agent (LIBRA) was developed. The data source used for this research mainly involved extract from Amazon.com web pages.

Gomez-Uribe and Hunt (2016) developed a system named as The netflix recommender system: Algorithms, business value, and innovation, ACM Transactions on Management Information System. The methods used in this are Top N algorithm, PVR (personalized video ranker) algorithm, supervised (classification, regression) and unsupervised approaches (dimensionality reduction through clustering or compression). It concludes that the different algorithms that make up the Netflix recommender system, the process that paper use to improve it. Paper convinced that recommender system will effectively guiding people to the truly best few options for them to be evaluated, resulting in better decisions.

Nath Nandi et al. (2018) used Doc2Vec algorithm for news recommendation system in Bangla language. It concludes

Doc2vec performs better than two popular topic modelling techniques LDA and LSA.

Lund and Ng (2018) proposed a system as Movie recommendations using the deep learning approach. It uses Collaborative filtering – k nearest neighbour, matrix factorization. Results shows that the recommendation system outperforms a user based neighbourhood baseline both in terms of root mean squared error on predicted ratings.

After taking the overview of the above researches, the specialized and personalized system for recommending the perfumes has not developed yet. While searching any perfume we are interested, we can enter the description of what we exactly need. To query these systems we can use our likes, dislikes i.e. sentiments about the perfume to get relevant recommendations. Hence, considering use of perfumes recommendation system will help users to get the perfumes which meet their requirements.

III. METHODOLOGY

The aim of this research is to build a perfume recommendation system. This system will help the user to get required perfumes. For that user has to provide description as a search query about the perfume according to his interest. This description can contains feelings, emotions, description, likes, dislikes and brand of the perfume. A chat bot will help the user to get the input in the form of search query and then provide the output as a recommended perfumes what user is looking for.

Initial work for research is collecting a data. Data required for this research contained the details in the form of name, brand, text descriptions, reviews, a list of notes.
As we are using natural language processing, the text data must be preprocessed. It covers some tasks like making text data to lower case, removing stop words, tokenization, stemming, etc. The Fig.1 shows tasks of preprocessing of data.
Lowercasing – Lowercasing is the first step in data
        preprocessing. The step is simple but important. Making the entire text to lowercase is essential step to get the consistent output. lower() method returns the lowercased string from the given string. It converts all uppercase characters to lowercase.
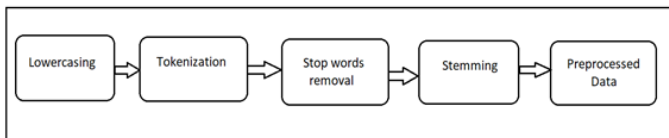

Fig 1: Data Preprocessing

If no uppercase characters exist, it returns the original string.
Tokenization – Once lowercasing is done, tokenization will take place. It means the sentences are divided into substrings known as Tokens. These tokens can be used to find the words in sentences. Tokenization basically refers to splitting up a larger body of text into smaller lines, words. The various
        tokenization functions in-built into the nltk module itself. A simple regular expression based tokenizer RegexpTokenizer provided by NLTK was used which splits the text into punctuations and whitespaces.
Stopwords Removal - Stopwords are the English words which does not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the sentence. For example, the words like the, he, a, an, have etc. Once the stop

words get removed we can focus on important, meaningful words. A Natural language toolkit (NLTK) was used to load the stopwords and remove them.
Stemming - Stemming is the process of bringing word to their root form. Stemming is the process of producing morphological variants of a root/base word. Stemming programs are commonly referred to as stemming algorithms or stemmers. A stemming algorithm reduces the words "chocolates", "chocolatey", "choco" to the root word, "chocolate" and "retrieval", "retrieved", "retrieves" reduce to the stem "retrieve". A SnowballStemmer provided by NLTK was used to perform stemming on the
        dataframe.

IV MODELING

This paper uses natural language processing approach to recommend the relevant product to the customer. It uses 2 models named as : Doc2Vec (Paragraph Vectoring) and Latent Semantic Analysis (LSA). These models are simple to use and provide high accuracy. The accuracy of the LSA to recommend the product is 84% and that for Doc2Vec is 91%. These models were evaluated against cosine similarity and individual similarities of the models were averaged to get accurate and better recommendations.

Latent Semantic Analysis (LSA) :

All languages have their own intricacies and nuances which are quite difficult for a machine to capture. This can include different words that mean the same thing, and also the words which have the same spelling but different meanings. A machine would not be able to capture this concept as it cannot understand the context in which the words have been used. Then, Latent Semantic Analysis (LSA) comes into play as it
        attempts to leverage the context around the words to capture the hidden concepts, also known as topics.LSA (Latent Semantic Analysis) also known as LSI (Latent Semantic Index) LSA uses bag of word (BoW) model, which results in a term-document matrix (occurrence of terms in a document). Rows represent terms and columns represent documents. LSA learns latent topics by performing a matrix decomposition on the document-term matrix using Singular value decomposition. LSA is typically used as a dimension reduction or noise reducing technique.

Singular Value Decomposition (SVD) - SVD is a matrix factorization method that represents a matrix in the product of two matrices. It offers various useful applications in signal processing, psychology, sociology, climate, and atmospheric science, statistics and astronomy.

$N = VEW^*$

Where, N is m*m matrix

V is m*n left singular matrix

E is n* n diagonal matrix with non-negative real numbers

W is m*n right singular matrix

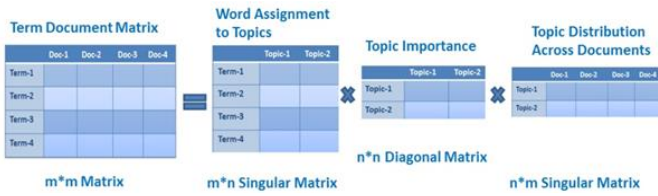W* is n*m matrix, which is transport of W

Fig 2: LSA

Doc2Vec (Paragraph Vectoring) :

Paragraph vectoring or Doc2Vec is an unsupervised learning algorithm that generates a fixed-length representation of a variable-length text document. The goal of doc2vec is to create a numeric representation of a document, regardless of its length. But unlike words, documents do not come in logical structures such as words, so the another method has to be found. The concept that Mikilov and Le have used was simple, yet clever: they have used the word2vec model, and added another vector (Paragraph ID below). Instead of using just words to predict the next word, we also added another feature   vector, which is document-unique.

So, when training the word vectors W, the document vector D is trained as well, and in the end of training, it holds a numeric representation of the document. The model above is called *Distributed Memory version of Paragraph Vector* (PV-DM). It acts as a memory that remembers what is missing from the current context — or as the topic of the paragraph. While the word vectors represent the concept of a word, the document vector intends to represent the concept of a document. The doc2vec models may be used in the following way: for training, a set of documents is required. A word vector W is generated for each word, and a document vector D is generated for each document. The model also trains weights for a softmax hidden layer. In the inference stage, a new document may be presented, and all weights are fixed to calculate the document vector.
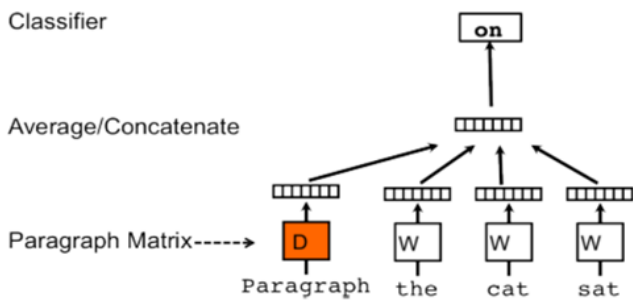

Fig 3: PV-DM model

Evaluation :

This model consists of two document embeddings, one from LSA and the other from Doc2Vec. To train the LSA and Doc2Vec models, we will concatenate perfume descriptions, reviews, notes into one document per perfume. We will then use cosine similarity to find perfumes that are similar to the positive and neutral sentences from the chatbot message query. I will remove recommendations of perfumes that are similar to the negative sentences. Based on Cosine similarity, the details of the perfume and the query will be matched to recommend the perfumes to the user. Output i.e.

recommended perfumes are displayed with the help of chat bot interface.

V   DESIGN SPECIFICATION

To develop an efficient recommendation system that would help in generating recommendations successfully, an architecture design shown in the figure Figure 5. System is divided into three blocks. The first block refers the chat bot interface. It helps user to communicate with the system. Customer provide information about perfumes what he/she wants to buy. Customer can be as detailed as he/she likes. Second block named as Data Persistence Layer, which contains data source and data cleaning. Business Logic Layer contains pre-processing of data, document generation, Training and evaluation of models using Latent Sentiment Analysis (LSA) and Doc2Vec.
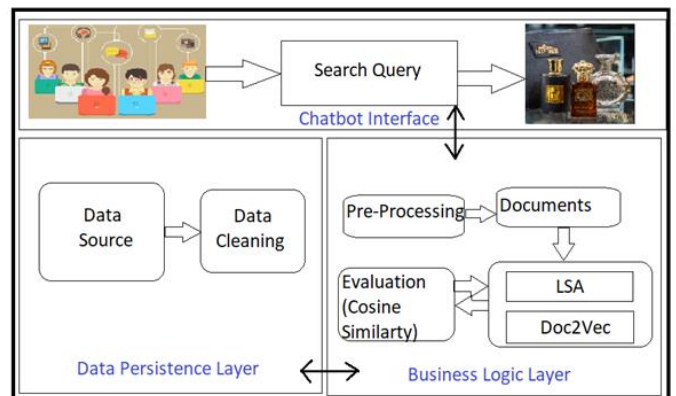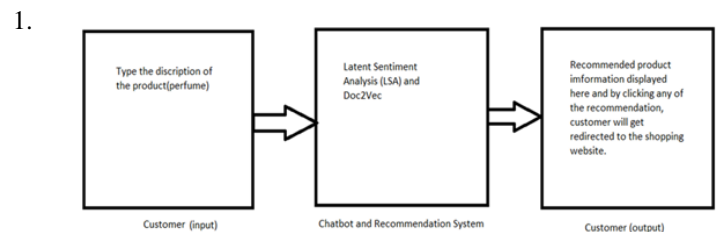

Fig 4 : Architecture

A. Flow Diagram

1.

Fig 5 : Flow Diagram

2. System Flow mentioned in Figure 6. Customer view is basically the chat bot interface from which user can provide input in terms of description of perfume which user wants to buy. Next block evaluates the result using LSA, Doc2Vec and cosine similarity. Last Block presents the output which shows the recommended perfumes according to user's description.

VI CONCLUSION AND FUTURE WORK

3. This research built a recommendation system model for perfumes using the dataset of niche perfume. When user provide description about perfumes according to his interest, system will recommend the relevant perfumes. We have found that combining the models: LSA (Latent Semantic Analysis) and Doc2Vec (Paragraph Vectoring) is more helpful to get the accurate results. By using this two methods and above approach we can say that system will provide more relevant results. Thus, the use of chat bot makes the system more interactive. Users provide the input in the form of description of perfumes according to their likings, needs, moods, events, etc. System will provide relevant products using LSA and Doc2Vec embeddings.

This type of a model can be used for any specific product such as books, electric appliances, musical instruments, clothing, shoes, watches, cars etc. The significant benefit of this type of recommendation system is, it worked on sentiment and semantic of users and then provide relevant results. This type of system deployed in a live environment, in the near future, would generate business and be more competitive in today's world of e-commerce.

VII ACKNOWLEDGEMENT

REFERENCES

[1] Bobadilla, J., Ortega, F., Hernando, A. and Guti´eRrez, A. (2013). Recommender systems survey, Know.-Based Syst. 46: 109–132.

[2] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. R. H. and Wirth, R. B. (2000). Crisp-dm 1.0: Step-by-step data mining guide.

[3] Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B. and Sampath, D. (2010). The youtube video recommendation system, Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10, ACM, New York, NY, USA, pp. 293–296.

[4] Gomez-Uribe, C. A. and Hunt, N. (2016). The netflix recommender system: Algorithms, business value, and innovation, ACM Transactions on Management Information Systems (TMIS) 6(4): 13.

[5] Han Lau, J. and Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation, pp. 78–86.

[6] Hutto, C. and Gilbert, E. (2015). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Isinkaye, F., Folajimi, Y. and Ojokoh, B. (2015).

[7] Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal 16(3): 261–273.

[8] Kucherbaev, P., Psyllidis, A. and Bozzon, A. (2017). Chatbots as conversational recommender systems in urban contexts, Proceedings of the International Workshop on Recommender Systems for Citizens, CitRec '17, ACM, New York, NY, USA, pp. 6:1– 6:2.

[9] Lahitani, A. R., Permanasari, A. E. and Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment, 2016 4th International Conference on Cyber and IT Service Management, pp. 1–6.

[10] Lund, J. and Ng, Y. (2018). Movie recommendations using the deep learning approach, 2018 IEEE International Conference on Information Reuse and Integration (IRI), pp. 47–54.

[11] Ma, H., Wang, X., Hou, J. and Lu, Y. (2017). Course recommendation based on semantic similarity analysis, 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), pp. 638–641.

[12] Mooney, R. J. and Roy, L. (2000). Content-based book recommending using learning for text categorization, Proceedings of the fifth ACM conference on Digital libraries, ACM, pp. 195–204.

[13] Nath Nandi, R., Arefin Zaman, M. M., Al Muntasir, T., Hosain Sumit, S., Sourov, T. and Jamil-Ur Rahman, M. (2018). Bangla news recommendation using doc2vec, 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), pp. 1– 5.

[14] Pan, C. and Li, W. (2010). Research paper recommendation with topic analysis, 2010 International Conference On Computer Design and Applications, Vol. 4, pp. V4–264– V4–268.

[15] Schafer, J. B., Konstan, J. and Riedl, J. (1999a). Recommender systems in e-commerce, Proceedings of the 1st ACM conference on Electronic commerce, ACM, pp. 158–166.

[16] Tan, H., Guo, J. and Li, Y. (2008). E-learning recommendation system, 2008 International Conference on Computer Science and Software Engineering, Vol. 5, pp. 430–433.