Context-Aware Hotel Recommendation System based on Hybrid Approach to Mitigate Cold-Start-Problem

Khushbu Jalan

Computer Engineering Dept.
Sardar Patel Institute of Technology (SPIT)
Mumbai, India

Prof. Kiran Gawande
Computer Engineering Dept.
Sardar Patel Institute of Technology (SPIT)
Mumbai, India

Abstract—Recommender System is an efficient tool that consider an individual's opinion to identify their content more appropriately and selectively. This system has been applied to various domain, but in the field of research area, service based recommendation system plays a major role. In recent years, the growth rate of online hotel searching has been increased much faster and makes this online hotel searching a very difficult task due to the abundant amount of online information. Reviews written by the travelers replaces the word-of-mouth but then to searching becomes the time consuming task based on user preference. Reviews crawled from the travelers visiting sites are a common and valuable source of information for recommendation of a hotel, yet little attention has been paid as, how to present the reviews of a reviewers in an understandable format.

The purpose of this project is to recommend the travelers the name of the hotels based on their preferences, by analyzing the other travelers reviews together with the rating value to improve the prediction accuracy. Users do not rate adequate hotels to enable collaborating filtering based recommendation, which can lead to an issue called as cold start problem. To solve this issue context-aware hybrid approach is used where collaborative filtering technique aggregates with the sentimental analysis, to provide personalized hotel recommendation. Here aspect based sentiment analysis is performed where weights of each aspect is calculated to find its orientation score. Also we use semisupervised clustering algorithm to group the aspects of similar meaning. Lexicon based approach is used to identify the sentiment towards the hotel aspects based on the defined context. Thus, context-based approach increases the recommendation results.

Keywords—Recommender system, Aspect based Sentiment Analysis, Cold start problem, Context based Recommendation system

I. INTRODUCTION

As Web 2.0 internet has been increasing, the number of user's using the web applications has been increased. Internet has large amount of information about peoples expressed reviews and opinion, which makes it as an excellent place for the source of information to gather data about a particular

object in a particular domain. Nowadays, products and services offered by most of the social web-sites are accompanied by abundant user-generated reviews, which can help users in decision making process. Decision makers extract the useful data from the user generated reviews which helps them to take the decision in a particular domain. Mining valuable opinion information from the plentiful but unstructured and noisy, source of user's content help analyze user's behavior which further helps recommendation task. Recommendation system is the best information filtering technique giving valuable information to user from user generated content. Recommender systems provides the user preference which helps the user to make the choice or examine it. Information used for recommendations can come from different sources like browsing and searching data, purchase data, feed-backs explicitly provided by the users and textual comments.

One of the renowned application making use of a recommender system is a Hotel Industry. Hotel which suits ones requirements in terms of price, service, quality, location and others is very complex task. Recommendation Systems play a major role by filtering the places which suit the users interests and requirements. Traditional Recommendation system only utilizes user's rating history. Therefore recommendation for travel is quite difficult, as user hardly rate enough hotels which makes difficult to build user profile. Having limited or no information about the user leads to the cold-start-problem. Considering Context-aware Recommendation system increases the accuracy by extracting the contextual information from user's review, like traveler type such as "Business type", "Couple", "Solo", "Family" and "Group", user's location and user's preferences for different hotel aspects/features. For example: "Wifi" can be an important feature for the user who is on the Business trip whereas for a Family trip, this feature may not be important. "Gym/Pool" can score a high rating feature for the family trip. **Table-I** shows the example of context-based travelers reviews consisting of features and sentiments. Based on user preference, our recommendation system mines the features context wise and determines its orientation related to hotels

TABLE I. AN EXAMPLE OF A CONTEXT-BASED TRAVELERS REVIEWS

1	Beautiful place to stay with Family, nice staff and roof top restaurant. With swimming pool and nice interiors - feels like staying in old times Heritage Mahal. Well made and maintained. Good and supportive staff
2	Cot to stay at this hotal, while on a business trin

Got to stay at this hotel, while on a business trip.
The hotel offers you a perfect blend of hospitality and comfort. The Food stands for a class of its own.

presented in the reviews. Main goal is to design a Hybrid based Recommendation System based on user's context, ratings and their reviews. Mainly in recommendation system, hybrid approach combines content based filtering and collaborative filtering but in this paper we are alternating the hybrid approach where recommendation results will be more improved by incorporating sentiment analysis with collaborative filtering approach. Cold-start-problem can be solved and gives high performance recommendation result by using context based information and applying hybrid based technique.

In our paper we define three types of context information. First is type of traveler which itself is divided into five categories like business type, couple, individual, group and family. Second one is user's location and third one is user's preference on different hotel aspects. We then mines this context based feature using text mining technique. Weights of an extracted features are calculated which further is used to get the overall sentiment score of the reviews. Here we use clustering algorithm to group the feature related words that mentions each hotel aspects from the text. We extract the different aspects of hotel and user's opinions from the reviews and groups the representative aspects by similarity to generate an output summary.

We obtained the data from the Tripadvisor.com. The details of each hotel is stored in a database, which includes hotel general information, it's reviews and ratings. Using the context based information for hotel searching gives more accurate recommendation.

- Collaborative Filtering Technique(CF): It is a standard recommender system technique that takes the preference information from many users. A technique that predicts the interest of a user by collaborating with many other users and finds a subset of users who have similar tastes and preferences to the target user and use this subset for offering recommendations. The technique[18] uses the numerical values given by the user in an online information. CF are of two types: (1) Memory based filtering: this method use the user's rating data to figure out the similarities between the users. (2) Model based filtering: this method use the data mining and machine learning technique to predict the user's rating which has not rated adequately to provide recommendation. This method can be used to mitigate issues of CF like cold start problem, data sparsity, etc.
- Sentimental analysis: Sentimental analysis also specified as opinion mining, analyzes the user's

opinions, attitudes, emotions, sentiments and evaluations from the online reviews [19]. Mainly the most important pointers of sentiments are sentiment words, that classifies the information into positive or negative sentiments. For example, good, best, superb, and exceptional are positive sentiment words, and bad, worst, terrible and horrible are negative sentiment words. Apart from individual words, there are also phrases and idioms eg. "The room service was great". To achieve the goal of mining opinions, task done by sentimental analysis technique is determine the orientation from the reviews and also to find the strength of the same.

Rest of the paper is organized as follows: The next section deals with the motivation and reviews the existing literature followed by the explanation of proposed system and lastly analyzes the result which concludes the paper with recommendation on future work.

II. MOTIVATION AND RELATED WORK

A. Motivation

As we have studied earlier, recommender system can recommends the possible related items to a user. Recommending any products or services to a user has become an essential application with the growing technology. Text based recommendation system is developing with the tremendous speed in many research areas, with the goal to include the valuable source of information from usergenerated content into the recommending process. As only ratings were considered in traditional recommendation system, which leads to several problems like scalability, data sparsity, cold start user etc. To solve the issue of cold start user, Hybrid Recommendation approach is undertaken taking Hotel Industry as a domain where hotel reviews and ratings are considered along with user's opinions and context that enhances the prediction using hybrid method.

B. Related Work

User behavior can be analyzed by exploiting the user generated contents available abundantly on the internet. Recommendation system[1] is an efficient tool that manages to filter this abundant information according to the user's preference and interest. Traditional recommendation system uses Collaborative filtering technique to recommend any products or services. But the main issue faced by this technique is cold start problem which is defined as cold start user or cold start item[2], and has been captured by many researchers i.e where no rating information is yielded when a new user register the system or not much information is found about the new item to recommend. This gives rise to different methods and techniques associated with CF technique. Hybrid approach, where CF technique integrates with the sentiment analysis tries to resolve the issue of cold-start-problem making use of contextual information from travelers textual reviews, thus resulting a better performance.

Khan, Khairullah et al. in [3] gives the detail study about the opinion mining. The study deals with subjectivity analysis and lexical resource generation. This study exploits opinion representation, opinion mining models and opinion components by retrieving opinions from primary sources like social networks and web blogs.

G Adomavicius et al. in [4] shows how contextual information added importance in providing recommendations. Three algorithmic patterns contextual pre-filtering, post-filtering and modeling were introduced to boost the contextual information in recommendation process. Also a case study is presented describing an unified approach of contextual information methods and techniques.

Jannach et al. in [5] compares different models of regression and have constituted aggregation functions which analyzes differently with respect to users and items. They use SVM and conclude that a support vector machine will perform better than linear least squares regression models if used for learning a combination of user and item. They also constitute that when weights are given, specific regression models performs the best.

Duan et al. in [6] presents how to mine the online user reviews that is used as both quantitative aspects and textual content from multi-dimensional perspectives. Online user generated reviews for the hotel industry used as data source, which by using a sentiment analysis technique decomposes the reviews into five dimensions for measuring and capturing service quality dimensions. This dimensions are then assimilated into econometrics models to get the overall evaluation of the content generating behaviour. Through sentimental analysis a notation of personal scale, based on the observation that different users gives different ratings to their reviews to improve the polarity classification analysis, which provides personal recommendation is explained in [7]. Text reviews are used to build a user profile, but cold start problem is not concentrated.

Gang et al. in [8] proposed an emerging pattern mining technique where the hotel managers analyzes the travelers concern from emerging hotel features w.r.t response to improve their strategic planning, marketing and product development.

Ye Hongwu et al. in [9] proposed a method for finding nearest neighbour through self-organizing-map which makes a group of nearest neighbours which is first step in collaborative filtering. Association mining is used to fill vacant space. Thus they proposed combination of association mining and SOM to address the issue of data sparsity.

Wu Yao et al. in [10] proposed probilistic model FLAME: was introduced addressing the problem of Personalized Latent Aspect Rating Analysis. A unified probabilistic model called Factorized Latent Aspect Model (FLAME), explains the advantages of collaborative filtering and aspect based opinion mining, combination of both solves the problem of latent aspect rating. Probilistic model learns the user's personalized preferences on different aspects from their past reviews, and predicts user's aspect ratings on new items by collective intelligence.

AK Samha et al. in [11] focus to improve the accuracy of extraction mainly by combining the different techniques from

3 different areas, i.e namedData Mining, Natural Language Processing and Ontology and generate the aspect-based opinionated summary from the customer reviews. Opinion mining for the application of hotel searching, where reviews are mined on the basis of combination of two Naive Bayes classifiers is proposed in [12]. Classifiers have different feature selection techniques: Relief-F and Chi-square. Ensemble model was compared with standard searching methods: Boolean and Boyer-Moore. Ensemble model increases the searching speed and provides highest average rank-accuracy.

HN Kim et al. in [13] In this paper a unique method is proposed of designing error reflected models obtained from explicit ratings. First actual rating is predicted, then based on this model identifies the predicted error for each user.

Li, Yize, et al. in [14] proposes a novel method to solve the issue of cold-start-problem and extraction of context from reviews to increase the recommendation quality. A probalistic latent relation model is introduced to integrate contextual information with recommendation algorithms.

FZ Lahlou et al. in [15] presents the different text classification techniques i.e, Naive Bayes, Decision Tree, k-Nearest Neighbours and Support Vector Machines to extract the context for context recommendation. The technique is related to stemming, minimum term frequency thresholding, weighting schema and n-gram words. Further the best technique is identified giving best classification.

A Muangon et al. in [16] proposed a framework of feature based opinion mining that makes use of two main lexicon features and polar words. Syntactic pattern analysis is used to extract the lexicons from the text and based on evaluated scores and lexical analysis, gives top polar words from reviews.

CWK Leung et al. in [17] describes a rating inference approach to include the text reviews in Collaborative Filtering algorithms. A relative-frequency-based method is used determine he sentiment orientation and strength of opinion words and then aggregate this orientation to conclude the average sentiment used by user. Paper also solves an issue of data sparsity by integrating textual reviews into CF algorithms.

Despite of the much study on context-aware recommendation system, most information retrieval system is based on query collections while browsing contextual text ignoring the textual content having opinions in the usergenerated reviews. Our study aims to make personalized hotel recommendations by analyzing the user's behaviour from the hotel reviews which identifies the user preferred features like room, food, service, etc based on context for selecting the hotel.

III. SYSTEM OVERVIEW

Core idea of the proposed system is to give more importance to the travelers reviews categorizing in the same context. Aspect based sentiment analysis is focused which is a information retrieval method that extracts the opinions on aspects of entities from textual content. Once the user specifies his mode of intent, his location and preferences, opinions on aspects are evaluated context wise and recommendation of hotel is presented accordingly. Fig 1. shows the following phase of the system.

A. Data Collection Phase

First the datasets were collected from Tripadvisor.com and stored in a database. Database contain hotel's general information ,user's reviews and overall rating of an hotel as well as aspect based rating.

B. Preprocessing Phase

- Sentence Tokenization: Every sentence is split into tokens on white spaces and punctuation marks. These tokens are usually words, but can also be numbers or symbols. Usually all whitespace characters are removed from the sentences and alphabets or numbers are considered as a single token if they don't have a whitespace or a special character like a tab-space in between.
- Stop Word Removal: Words such as 'a', 'about', 'also', 'and', 'at', 'to' etc. are removed. There is no single, universally approved, list of stop words present in the English language.
- Word Replacement: "The food was gr8" is classified as neutral classification whereas "The food was great" comes under positive sentiment classification. Because this is the world of short messages we have created a dictionary of some words having short meanings and stored in a database. While processing the corpus of reviews the words are corrected before we take the reviews as an input.
- Stemming: Porter Stemmer algorithm is used to remove the common morphological and inflexional ending words from English dictionary. It is a kind of normalization process that is used for information retrieval system.
- Parts-Of-Speech Tagging(POS): Each word in the corpus is labelled as to which part of speech it belongs to. This is done by parsing the reviews and extracting the important information from the reviews such as features and opinions. POS tagger parse the sentence it lies in, the words which are adjacent to the word being tagged, the phrase and the paragraph. We used OpenNLP tools to parse the sentences. The words are tagged as Singular Nouns, Plural Nouns, Verbs, Adjectives, Adverbs etc. The tagging is done using the Stanford core NLP which returns tags like NN for singular common nouns, NP: singular proper nouns etc. Also we used a WordNet database to find the synonyms to extend our aspect list.
- Feature Extraction: To extract the features we need to analyze the reviews to know, POS tagger has tagged which word as features and which word as a opinions. Like nouns are identified as features. For example

"Location", "Service", are the features of hotel. Here the features are extracted at the sentence level.

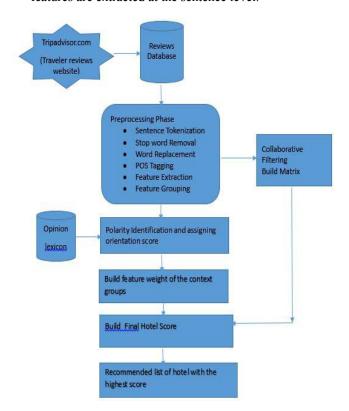


Fig. 1. Proposed Flow Diagram

Aspect Grouping: semisupervised clustering algorithm
is used to process the corpus of reviews to find the
vocabulary that is used to describe the particular
aspect/feature of a hotel with the same meaning and
cluster the similar one into one group[20]. Aspect
grouping is difficult due to various possible synonyms.
Difficulties like words referring to the same aspects is
not considered as synonyms in some dictionary and
many synonyms are domain specific. Table-II describes
the features of hotel with their synonyms.

C. Polarity Identification

Polarity of sentiment(Positive and Negative) on each aspect is determined by lexicon based approach. It's a dictionary of words and word phrases that classifies words into positive and negative sentiments. All adjectives are extracted from the sentences using this approach for each feature. Now an orientation score is assigned to each feature fe in a given sentence se, denoted as score(fe,se). It should be clear that the same feature, in two different sentences, could receive different orientations. When many opinion words surround a single feature, they are aggregated as:

Score(fe, se)=
$$\sum_{op \in s} \frac{o \, r_{op}}{d \, (op, fe)} \tag{1}$$

TABLE II. ASPECT GROUPING USING SEMISUPERVISED CLUSTERING ALGORITHM

SR. No.	FEATURES	LIST OF FEATURES
1	FOOD AND DRINK	Coffee shop, pool-side bar, breakfast, morning, food, restaurant, bar, coffee, buffet,dinner, fruit, terrace, buffet variety, bread, course
2	LOCATION	location, area, city, street, metro, station, train, distance, bus, airport
3	SERVICE	Staff, Reception, Laundry, Dry cleaning, Cash withdrawal, Smoking area, service, hotel staff, front desk, luggage lobby, luggage staff, person
4	ROOM COMFORT & QUALITY	bathroom, floor, shower, size, window, door, view,building, water ,balcony, hotel room, bath, AC, Tv
5	ACCESSIBILITY	elevator, lift, disabled guests, wheel chair, chk-in, chk-out(express), front-desk.
6	PRICE/VALUE	price, cost, rate, money value, money, value
7	CLEANLINESS	clean
8	FACILITIES	pool, spa, sauna, steam room, gym, games room, wifi
9	GENERAL	hotel, night, place, stay, experience, trip, hotel star, deal, quality

In (1) op is the opinion word in the sentence se, d(op,se) is the distance between the feature and opinion word in the sentence. Also, or_{op} is the orientation (1, +1) of the opinion word op.

D. Build Feature Weight

We build a feature weight based on the context groups by counting the number of times the particular feature occurs. For the n^{th} aspect, let A_n be the total occurrences of feature words in the n^{th} aspect. The weight of each aspect is defined as:

$$W_n = \frac{A_n}{\sum_{n=1}^{num} A_n}$$
 (2)

where num denotes the total number of identified aspects. Therefore for each context groups weights will be defined as W_{tt} for context traveler type, W_{loc} for user's location and W_p for user's preference.

E. Build Final Hotel Score

Subsequently combine the features, their weights and orientations as per the user's context to build a score for each sentence. The sentence scores are then combined to give an overall score for each review. The hotels with higher orientation scores based on features will be recommended to the user. Let us calculate the sentiments on review r for a particular aspect A_n that are based on all the sentences of the online reviews. The equation of aggregated sentiments is defined as:

$$sen_r(r, a_n) = AVG_{fe \in a_n, se \in r} (sen_se(fe, se))$$
(3)

By averaging the sentiments calculated above from the reviews, we get the overall sentiments of hotel ht which is defined as:

$$sen_ht(ht, a_n) = AVG_{r \in ht}(sen_r(r, a_n))$$
(4)

Finally the total recommendation score of a hotel ht is calculated based on user's selected features context wise.

$$Score(ht) = \sum_{n=1}^{num} W_n sen_ht(ht, a_n)$$
(5)

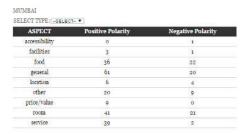
Hotel with the high score is ranked accordingly and recommended.

IV. RESULT ANALYSIS

We retrieve the reviews of at least 5 hotels from Tripadvisor.com to evaluate the proposed hotel recommendation approach. Mainly each hotel consists of hotel's general information such as(hotel id, user name, address) and reviews plus overall rating. The reviews basically contain the contextual information that is mainly focused i.e traveler type, user's location, feature rating and other additional information. Now based on each context, the feature is extracted which forms a lexicon. Based on feature, orientation of feature is identified at a sentence-level.

The figure-2 below shows the total features extracted from the reviews. Further, aspect based sentiment is identified using lexicon based approach and total positive and negative score is calculated as shown in fig-2.

WELCOME TO HOTEL RECOMMENDATION SYSTEM



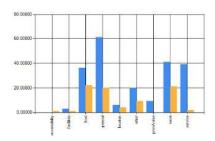


Fig. 2. Graph showing feature based orientation along with the score

Figure-3 shows that as the user context is defined, feature orientation score is calculated accordingly as per user's preference.

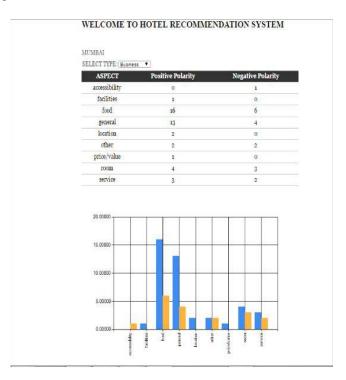


Fig. 3. Graph shows feature-based orientation based on context Business type

Figure-4, shows that based on orientation score, top hotel is recommended context wise.

WELCOME TO HOTEL RECOMMENDATION SYSTEM



Fig. 4. Context based Hotel Recommendation

V. CONCLUSIONS AND FUTURE WORK

The paper is concluded with the main Objective of the project along with the system overview. Mainly project is designing a Recommendation System Based on Hybrid Approach. Approach is dependent on context based. Therefore context based Recommendation system is developed which

defines context group as type of traveler, user location and users amenities preference. This paper studies the first part of the project. First features are extracted using text mining where semi-supervised clustering algorithm is used to group the features of the same vocabulary. We have grouped the words of same meaning into nine categories which are our extracted features. Then, lexicon based approach is used to perform aspect based sentiment orientation. We have shown the total positive and negative sentiment score first for total corpus of reviews and then by user's given preference. Results shows the better performance in giving recommendation when context based reviews are considered. Moreover in future work we are going to aggregate this orientation score with the rating matrix to solve an issue of cold-start-user. Average rating of all the aspects will be calculated which integrates with the orientation score to give rank wise hotel recommendation.

References

- [1] Chen, Li, Guanliang Chen, and Feng Wang. "Recommender systems based on user reviews: the state of the art." User Modeling and User-Adapted Interaction 25.2 (2015).
- [2] Bobadilla JesuS, Fernando Ortega, Antonio Hernando, JesuS Bernal. "A collaborative filtering approach to mitigate the new user cold start problem." Knowledge-Based Systems 26 (2012).
- [3] Khan Khairullah, Baharum Baharudin, Aurnagzeb Khan, Ashraf Ullah. "Mining opinion components from unstructured reviews: A review." Journal of King Saud University-Computer and Information Sciences 26.3 (2014).
- [4] Adomavicius, Gediminas, Alexander Tuzhilin. "Context-aware recommender systems." Recommender systems handbook. Springer US, 2015.
- [5] Jannach D., Gedikli, F., Karakaya, Z., Juwig, O.: Recommending hotels based on multidimensional customer ratings. In: International Conference on Information and Communication Technologies in Tourism (ENTER), Springer (2012)
- [6] Duan Wenjing, Qing Cao, Yang Yu, Stuart Levy. "Mining online user-generated content: using sentiment analysis technique to study hotel service quality." System Sciences (HICSS), 2013 46th Hawaii International Conference on. IEEE, 2013.
- [7] Faridani, Siamak. "Using canonical correlation analysis for generalized sentiment analysis, product recommendation and search." Proceedings of the fth ACM conference on Recommender systems. ACM, 2011.
- [8] Li Gang, Rob Law, Huy Quan Vu, Jia Rong, Xinyuan Roy Zhao. "Identifying emerging hotel preferences using EmergingPattern Mining technique." Tourism management 46 (2015).
- [9] Ye, Hongwu. "A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map." JSW 6.4 (2011).
- [10] Wu, Yao, and Martin Ester. "Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering." Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. ACM, 2015.
- [11] Samha, Amani K., Yuefeng Li, and Jinglan Zhang. "Aspectbased opinion extraction from customer reviews." arXiv preprint arXiv:1404.1982 (2014).
- [12] Srisuan, Janasthar, and Anantaporn Hanskunatai. "The ensemble of Nave Bayes classifiers for hotel searching." Computer Science and Engineering Conference (ICSEC), 2014 International. IEEE, 2014.
- [13] Kim, Heung-Nam, Abdulmotaleb El-Saddik, and Geun-Sik Jo. "Collaborative error-reflected models for cold-start recommender systems." Decision Support Systems 51.3 (2011).

- [14] Li, Yize, Jiazhong Nie, Yi Zhang, Bingqing Wang, Baoshi Yan, and Fuliang Weng. "Contextual recommendation based on text mining." In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 692-700. Association for Computational Linguistics, 2010.
- [15] Lahlou, Fatima Zahra, Houda Benbrahimand, Asmaa Mountassir, and Ismail Kassou. "Context extraction from reviews for Context Aware Recommendation using Text Classification techniques." In Computer Systems and Applications (AICCSA), 2013 ACS International Conference on, pp. 1-4. IEEE, 2013.
- [16] Muangon, Ananchai, Sotarat Thammaboosadee, and Choochart Haruechaiyasak. "A lexiconizing framework of feature-based opinion mining in tourism industry." Digital Information and Communication Technology and it's Applications (DICTAP), 2014 Fourth International Conference on. IEEE, 2014.
- [17] Leung, Cane WK, Stephen CF Chan, and Fu-lai Chung. "Integrating collaborative filtering and sentiment analysis: A rating inference approach." Proceedings of the ECAI 2006 workshop on recommender systems. 2006.
- [18] Schafer, J. Ben, Dan Frankowski, Jon Herlocker, Shilad Sen. "Collaborative filtering recommender systems." The adaptive web (2007)
- [19] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends in Information Retrieval 2.12 (2008).
- [20] Dalal, Mita K., and Mukesh A. Zaveri. "Semisupervised learning based opinion summarization and classification for online product reviews." Applied Computational Intelligence and Soft Computing 2013 (2013).