

Prediction of Reviews Rating: A Survey of Methods, Techniques and Hybrid Architectures

Fouzi Harrag¹, Abdulmalik Als Salman², Alaa Alqahtani²

¹Computer Science Department
Farhat ABBAS University, Setif-1
Setif, 19000, Algeria
fouzi.harrag@univ-setif.dz

²Computer Science Department
King Saud University
Riyadh, Saudi Arabia
salman@ksu.edu.sa, alaa.s.alqhtani@gmail.com



ABSTRACT: *In the last recent years, the customer reviews has attracted an increasing interest by researchers and companies thanks to the advancement in E-commerce and business applications. The ability to successfully decide whether a review will be helpful to other customers and thus give the product more exposure is vital to companies that support these reviews, companies like Google, Amazon and Yelp! The problem is that some customers are giving rating in conflict with their comments obliging other reviewers to read the comments to understand the reason of this contradiction. Sentiment analysis is the process of automatically identifying opinions expressed in text on certain subjects. The accuracy of sentiment analysis has a direct effect on decision making in both business and government. Also, recommender system is one of the hot topics in the field of big data, especially with the development and growth of the World Wide Web.*

This paper presents a survey covering the methods in the field of reviews rating prediction. We also propose a hybrid architecture that combines techniques from sentiment analysis and recommender systems fields, to develop a solution that has the ability to predict rating of products while taking into consideration the analysis of sentiments included in the products reviews.

Subject Categories and Descriptors: H.3.1 [Content Analysis and Indexing]; H.2.8 [Database Applications]: Data mining;

General Terms: Sentiment Analysis, Opinion Mining, Recommender Systems

Keywords: Sentiment Analysis, Arabic Textual Reviews, Opinion Mining, Recommender Systems, Rating Prediction

Received: 21 October 2018, Revised 10 December 2018, Accepted 24 January 2019

Review Metrics: Review Scale- 0/6, Review Score- 4. 8/6, Inter-reviewer consistency- 78%

DOI: 10.6025/jdim/2019/17/3/164-178

1. Introduction

Nowadays, as we know the Internet is growing up thus the textual information also is growing very fast. One of this textual information is the customer comments or reviews. People usually prefer to read the reviews before buying or using a service to make the right decision. This behavior is also common before the existence of the Internet. From this amount of available data, researches attempt to handle and use these data to have a specific and useful knowledge. Sentiment analysis (SA) is the process of determining the opinion or feeling of a piece of text. Sentiment means feelings, attitudes, emotions and opinions. The applications of sentiment analysis are numerous such as politics or political science, law, e-commerce, sociology and psychology.

In e-commerce, the sentiment analysis is super useful for gaining insight into customer opinions; once they understand how the customer feels after analyzing their comments or reviews, they can identify what they liked and disliked and built things like recommendation sys

tems, or enhance the product or the service.

In general, people are influenced by others' opinions. As a real life example, a person will go and eat in a specific restaurant after asking the people who tried this restaurant before. This is a common behavior, and there are several statistical studies such as:

- A study states the influence of the reviews where 64% of them spent 10 minutes to read the reviews, and 33% spent half hour or more in reading online reviews. Also, before buying: 39% read around 8 reviews or more and 12% read 16 reviews or more [1].
- 90% of the buyers said: their buying decision is affected by the online reviews [2].
- The action after reading a positive review, 48% of the survey responders are motivated to visit the business' website, and 21% are shopping around [3].
- A survey shows how important is the reviews before buying products: where 44.8% said it is important and 35% said it is very important while 4.7% only said the reviews are unimportant [4].
- 84% of the survey responders said: they trusted the online reviews as personal recommendation (vs. 80% in 2015) [5].

1.1 Problem Statement

In the few recent years, Sentiment analysis has attracted an increasing interest. Sentiment analysis, also known as opinion mining, is the process of determining the attitude of textual opinions whether they are positive, negative or neutral. Such a process has different useful applications including analysis of product reviews, advertisements, customer satisfaction and bloggers' attitude about an event and so on.

Also, recommender systems are one of the hot topics in the field of big data, especially with the development and growth of the World Wide Web. There are a lot of researches done to improve the recommendation process. In business and marketing sector, usually, our decision is influenced by the product or the service reviews and other people's opinions. The higher star ratings lead to more sales and orders. We opt to combine between recommendation and sentiment analysis to predict products' ratings which help in buying decision.

The majority of researches in this field focus on English products' reviews, however, the studies and researches about the sentiment analysis of Arabic reviews are few. That is why the development of such system is very important for the Arabic community.

The aim of this study is to present a state of the art of methods and techniques in the field of reviews rating prediction. The study is also supported by proposing archi-

tecture of a hybrid system combining opinion mining techniques and algorithms to extract and analyze the sentiments from customer's Arabic reviews.

2. Background in the field of Reviews Rating Prediction

In this Section, background information will be presented such as Sentiment Analysis (SA) and recommender systems (RS).

2.1 Sentiment Analysis

Sentiment analysis (SA), also called opinion mining, as defined in [6] "is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes".

While the sentiment analysis studies and focuses on the opinions, the term opinion relies on the attitude and associated information by the person who wrote this opinion (opinion holder). And the term sentiment means the underlying positive or negative feeling implied by opinion [7]. In sentiment analysis, there are different levels of analysis: document level, sentence level and aspect level [8]. In document level, the document as a whole is expressing either a positive or negative. Usually, this level is not applicable where more fine-grained analysis is needed [7]. In sentence level, it determines the sentence opinion whether it is positive, negative or neutral [9]. In aspect level, it can find exactly what the people like and dislike [10]. In simple words, it targets the 'aspect' of the opinion for example, "iPhone 7 is incredible" so it is a positive statement that targets the iPhone 7 itself.

2.1.1 Sentiment Analysis Applications

Customer opinions about products and services are important to the organizations and businesses. Also, sentiment analysis has important applications in other areas and sectors. In the governments and the federal sector, they concern about the nation opinions and concerns, by analyzing them, it will help in the evolution of their services. Other governments may attempt to monitor the other countries interests and other issues through the social media, comment on the news web pages and so on. Sometimes this information is important in detecting the criminals, international relations and economics [11]. Sentiment analysis, in general, helps in understanding the sentiment and emotions about a particular phenomenon, entity or idea. So, it is useful in any of these situations where the opinion can be discussed as a text (it can be reviews, blogs, news, feedback, discussion or comments).

2.1.2 Sentiment Analysis Methods and Models

The sentiment analysis methods can be categorized as follows: Lexicon-based approach, Corpus-based approach and Hybrid [12]. In lexicon-based approach (unsupervised approach) the word or sentence sentiment is determined

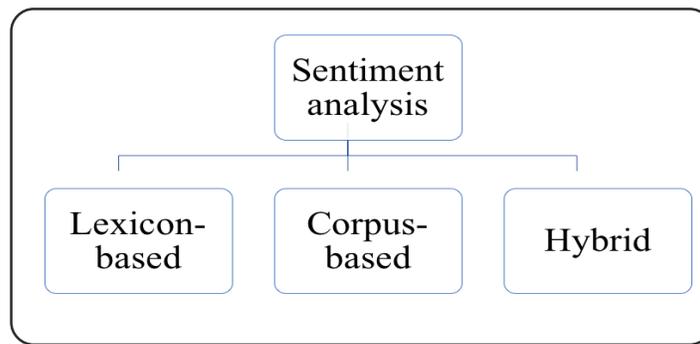


Figure 1. Sentiment analysis methods

using a lexicon [12]. In this lexicon, each word has a polarity value such as 1 for positive and 0 or -1 for the negative. And from that, the polarity of a sentence can be calculated. Figure 1 shows the classification of sentiment analysis approaches.

On the other hand, the corpus-based (supervised approach) one of the machine learning algorithms is used such as Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) where a manually annotated corpus is used in the training process of the classifier. After building the model, this model will be used in the classification process of the testing dataset [13]. The third method is the hybrid approach which is the combination of the above methods.

There are three sentiment analysis models: two-way classification (positive, negative), three-way classification (positive, negative, and neutral), and four-way classification (positive, negative, neutral, and mixed). Some of the researches proved the importance of the neutral class such as [14], where the authors used an Arabic dataset from Twitter to do their experiment. The obtained result shows the importance of having a neutral class with some specific classifiers, where a sentence or a word does not necessarily have a specific polarity (positive or negative).

2.1.3 Sentiment Analysis Challenges

Challenges in the sentiment analysis vary from one language to another since the sentiment analysis is a branch of Natural Language Processing (NLP). One of the sentiment analysis problems is detecting whether a statement has a polarity or not; some sentences do not express feelings such as “today, I want to have a good day”. Also, some of the sentences will be classified as neutral, while it does not contain any sentiment! Such as “Yeah, I will do it after I woke up” [15].

Another challenge is sarcasm, where the person writes something and the meaning does not meet what’s written in the statement such as ‘I see the diet is going well!’ in this clause, it seems the diet is fine while the meaning is totally the opposite. The researches that are handled or studied the sarcasm are few in English and most of the other languages, since it is hard to detect the sarcasm in the text. In addition, in some fields the sentiment could

be positive while in the other fields it is considered as negative, such as: ‘the sound of the washing machine is low’ and ‘the sound of the TV is low’ wherein the first comment is considered as an advantage for the washing machine if it has silence feature, while if the sound of the TV speakers is low it is a disadvantage!.

Also, as a challenge that may be faced by some application in the sentiment analysis is the identification of the entities in each sentence and identifying its features [16]. Sometimes, local dialects of the same language also will make the sentiment analysis process harder.

2.2 Recommender Systems

Recommender systems are filtering systems that show specific information to the user or customer and attempt to predict a rating or products [17]. It became popular in the recent years due to the increasing of the online shopping and transactions through the Internet. It is useful in many fields such as books, news articles, movies and restaurants.

Most of the websites with different areas allow the user to provide a feedback about what did they like and dislike. The most famous and oldest way of feedback is ratings; another kind is the textual notes or comments. The main idea of recommender systems is recommending an entity that belongs to the user interests [18]. Usually, the recommendation process depends on the user and items interactions, past behavior, relation to other users, item similarity or it can be a knowledge-based recommender system that relies on the user requirement rather than depending on the user past history [19].

2.2.1 Recommender System Applications

There are several applications of recommendation, such as the entertainment (e.g., movies, music), The most famous use is in the field of the e-commerce, where products are recommended to the customers. Another application is to recommend a service or software to the users. In addition, news personalization, web search, querying can be considered as application in this domain.

2.2.2 Recommender Systems Methods and Models

The main methods of recommender systems are dealing with two types, either the interaction between the user

and the item (such as rating) or information of the user and item (such as textual information). So, there are three methods: collaborative filtering, content-based method and knowledge-based method [20]. Some of these methods can be combined to introduce the hybrid recommender system. Figure 2 shows these recommender systems approaches.

The collaborative filtering methods are based on the information about the user such as user activities and behavior, collect and analyze this information to predict what this user may like and dislike based on the parity of other users, where the assumption is: the users who agreed in

the past will also agree in the future [21]. Content-based approach used the item features in the recommendation process where it assumes if the user will like items that similar to the liked item in the past [22]. Figure 3 and Figure 4 show the used matrix for these two methods, where some parts of these matrices are available and some are missing. Where the aim is finding and predicting these missing values based on the available information. Table 1 shows some of the advantages and disadvantages of each approach. In Figure 5, there is a comparison between collaborative filtering and content-based filtering in field of the e-commerce.

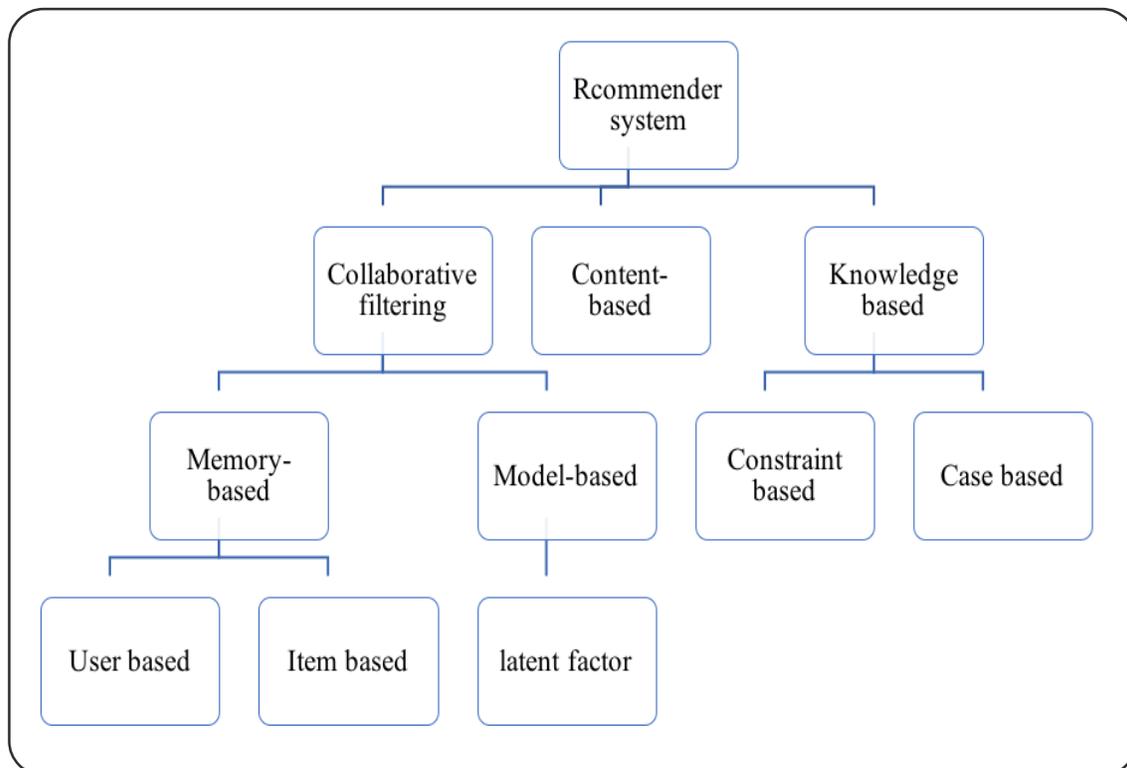


Figure 2. Recommender system methods

	items			
user	2		3	4
		1		
		3		2
	1		5	

Figure 3. Collaborative filtering: user-rating matrix

	features			
items	1			
		1	1	1
			1	
	1	1		1

Figure 4. Content-based filtering: item-content matrix

Knowledge-based approach is useful in case of the items that are not always purchased such as automobile or tourism request where there is little availability of ratings for the recommendation process. The most famous example is buying a house, where there is some constraint required such as the location, number of rooms and price. So, this method relies significantly on the items attributes. Knowledge-based is similar to content-based approach, but the main difference is the content-based based on user behavior in the past while in knowledge-based relies on the specific interests of the currently active user.

There are two main models: Prediction and ranking [23]. Where the prediction version aims to predict the rating of user-item set, this approach requires the availability of training data. The second approach is ranking, which does not predict the rating but the seller aims to show the top k-items to the customer as recommendations, and therefore the case of finding the top k-users is similar, but it is not common as the top k-items.

2.2.3 Recommender Systems Challenges

There are many challenges that faced collaborative filter-

Method	Advantage	Disadvantage
Collaborative filtering	<p>Good performance</p> <p>Requires minimal knowledge</p>	<p>Cold start problem: Not applicable if no enough ratings</p> <p>Requires large number of reliable feedback</p>
Content-based	<p>No need of items' ratings</p> <p>Ability of recommending new and un-popular items</p> <p>Explanations of recommended items can be provided</p>	<p>Ignore the users' relations</p> <p>Overfitting is easy</p>

Table 1. Advantages and disadvantages of recommender system methods

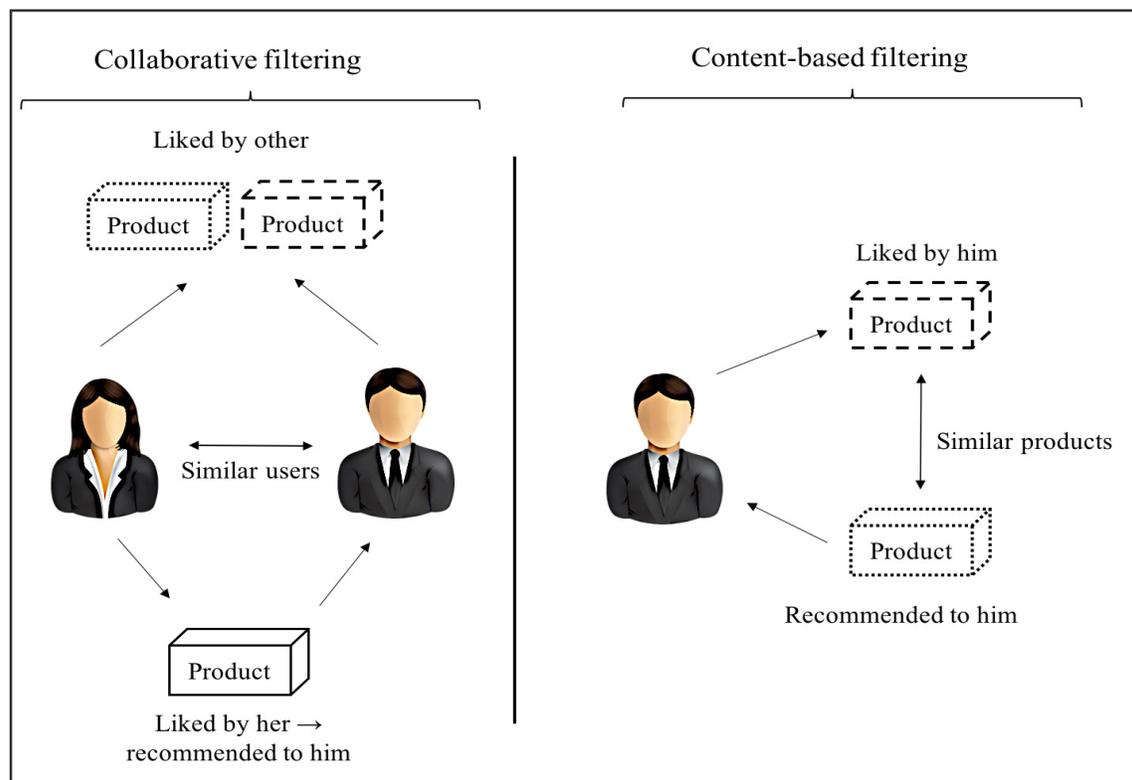


Figure 5. Comparison between collaborative filtering and content-based filtering

ing methods, such as the matrix sparsity [24]. Cold start problem another big challenge for collaborative filtering where the available ratings are not enough. Difficulty in making predictions that based on nearest neighbor algorithms [25] is also another known problem in this field. This problem is related to scalability, where the nearest neighbor requires computation that increases with both the users and items numbers.

On the other hand, content-based method also has challenges, where it requires content that could be encoded as meaningful features. Also, some kinds of items are not adjustable to easy feature extraction method (e.g., movies, music) [26]. User favorite items must be demonstrated as a learnable function of these items features. In addition, it is hard to achieve how much the quality and truth judgment of other users. The knowledge-based approach has similar challenges to content-based method.

3. Related works on Reviews Rating Prediction

The World Wide Web contains a huge amount of information that is easily accessible. This information is usually written as texts such as articles, blogs, reviews, comments and tweets. Recently, the researchers attempt to mine the opinions that expressed in texts to determine whether the attitude towards a particular topic, product, news and article or blog post is positive, negative or neutral. In this chapter, some of the related works will be presented.

In [27], the authors used the combination of sentiment analysis with information retrieval to predict the rating of Amazon comments. Vector Space Model (VSM) was applied as a supervised classifier. They compared it with the combination of VSM with sentiment analysis. The Lexical dictionary approach was used as sentiment analysis with the VSM. The obtained result shows that the usage of sentiment analysis has a positive effect on the performance of the classifier in their rating prediction.

While in [28], a subjectivity detection method was applied before classifying the polarity of student reviews written in the Arabic language. The subjectivity detection is the process that identifies whether a sentence contains an opinion or not. After performing this prior task, lexicon-based sentiment classification and Support Vector Machine (SVM) classifier were used to classify the attitude of the opinion sentences. At the end, the result shows that using subjectivity detection as prior task has better performance of sentiment classification since the error rate could be lowered and reduced when subjectivity detection is applied.

In [29], the main contribution was to generate sentiment Arabic lexical semantic database using lexicon-based approach; then, use it on sentence-level subjectivity and multi machine learning algorithms. Different types of classifiers were used including Bayes, Rules, Trees and Functions with 10 Cross-Fold-Validation. Using Functional

Trees (FT) as an algorithm for the sentiment analysis, the obtained result in terms of F-measure was 76.1% which was among the best result compared to the others. On the other hand, the generated lexicon can be used in sentiment analysis.

Moreover, in [30] also Arabic Sentiment Lexicon (ArSenL) was proposed. A large scale Arabic lexicon was created benefiting from existing Arabic lexica. At the beginning, two lexica were created, the first lexicon was generated by matching Arabic WordNet to ESWN and the second lexicon was generated by directly matching lemmas in SAMA to ESWN. So, the proposed lexicon ArSenL was the result of combining these two lexica by taking their union. The generated lexicon was publicly available, and for each lemma it has three scores to match the three sentiment labels: positive, negative and objective. ArSenL was evaluated by comparing the three generated lexica (first lexicon, second lexicon and the union) with the baseline and SIFAAT, and the result reflected that ArSenL "the union lexicon" has then better performance than SIFAAT.

Twitter (the famous social network) is rich in information and opinions. And since it has been used for sentiment analysis in many studies, hereafter are some of researches that apply sentiment analysis and opinion mining on tweets. In [31], the authors did an investigation on the customer satisfaction about telecom companies in Saudi Arabia, where Arabic tweets were collected. Semantic Sentiment Analysis (SSA) approach was used to analyze the collected tweets.

In addition, the authors in [32] studied the sentiment analysis of Arabic tweets. The collected tweets were written in Egyptian dialect. Two different machine learning algorithms were used, SVM and Naïve Bayes (NB). Firstly, they manually annotated the tweets into either positive and negative which were done by two raters and then revised by a third rater. Then, preprocessing was done to construct the feature vector. The obtained result shows SVM outperform NB, since NB is based on probabilities. Where in [33], authors proposed a framework that can detect the sentiment in Arabic tweets based on lexicon approach. In the first step, the lexicon was created by translating SentiStrength 'English sentiment lexicon' into Arabic. As unsupervised learning, this lexicon was used to determine the sentiment of the collected Arabic tweets. The overall sentiment of a tweet is the summation of the words' sentiment weights. The result shows that using lexicon improves the sentiment analysis. Also from the obtained result, stemming improved the overall accuracy.

While in [34], authors presented a system for subjectivity and sentiment analysis SAMAR for Arabic social media. The collected dataset is varied from different social media websites such as Twitter, Maktoob chat, Wikipedia talk and forums to mix between Modern Standard Arabic (MSA) and Dialectal Arabic (DA). This dataset was annotated by two native Arabic speakers. They divided their

work into two parts: subjectivity classification and sentiment classification. The obtained result reveals SAMAR outperform when part-of-speech (POS) used and combined with the standard features. And for sentiment classification result, also SAMAR improves the data compared to the baseline expect the data that are collected from forums. (In [35], they used a statistical software called R it is an open source tool. They used R's packages to harvest, preprocess, analyze and visualize tweets about the two famous stores in the United Kingdom: Tesco and Asda. The Lexicon-based approach was used. The lexicon that has been used in this study is publicly available. The constructed tool helps in tracking the user's opinions harvested from the social media such as Twitter, Facebook, and visualizes the result to the end user.

In another research [36], the authors attempted to analyze the type of Arabic language used by Saudi users in Twitter. This study focused on determining the preprocessing operations needed to be performed before any sentiment analysis in social networks data. According to their analysis on the collected data, the obtained results revealed that 80% of the collected tweets are written used Modern Standard Arabic (MSA). This study can help us to understand the famous Arabic language used in Saudi Arabia since we are trying to suggest a tool for our community.

The authors in [37], have considered both standard and dialectal Arabic. First, tweets were collected, filtered then preprocessed before starting the classification process using Arabic Sentiment Ontology (ASO). The obtained result showed that the use of the semantic approach could produce a good understanding of the overall image of some certain entities in the dataset.

In [38], the authors attempt to extract the aspect from the textual reviews. This will be helpful especially in the opinion mining, to help the customers to go directly and easier to the feature that he/she concerns about. For example 'this phone is amazing, it is lightweight and attractive' where the word 'lightweight' indicates the weight and 'attractive' points to the external appearance. They used the rule-based approach for that. They used two dataset which is publically available to compare their work with the previous work: first data set is the dataset provided by (Hu and Liu, 2004) and the second dataset is Semeval 2014 dataset. They divided their experiment into two categories based on these two datasets. The obtained result has higher accuracy in the aspect detection for both datasets.

In [39], proposed a document level (review-level) sentiment classification using the neural networks. They considered both the user and product information to achieve higher accuracy than considering one of them, which is one of the main contributions in this paper. They proposed User Product Neural Network (UPNN) model which take the user and product parameters as an input, and classify this review as an output (assign the polarity label to

it). They used three datasets in their evaluation: IMDB, Yelp 2013 and Yelp 2014. They compared their proposed model with the previous works that used the neural network in sentiment classification. The obtained result outperformed the state of artworks.

The traditional way in recommendation and recommender systems used factors like the category of the product, location and the record of the purchase. And in recent years the recommender system is not limited to the typical numerical information such as rating and user or item features, and due to the availability of valuable information in the textual format, we can use this information in the recommendation process. There are some research works that rely on the reviews in the recommendation. In [40], they proposed a bag of opinions method to use it in the review rating prediction. They used the textual patterns (product reviews) in their model to conclude the numeric rating of users. Also, they model a regression (constrained ridge regression) to calculate the numerical opinions scores from the review text. The dataset that used is collected from Amazon and they considered it as independent reviews (not in aa specific domain) from different categories books, movies and music, and then the model learned and then applied to learn the prediction in the domain specific such as movies reviews (e.g. in IMDB). Each review may include these three components: the root word (e.g. useful), modifier word (e.g. very, too) and negation words (e.g. not). Their result shows the mean squared error for their model is less than the compared works.

In [41], the authors proposed a rating predictor that depends on both: the sentiment of the product/service review and the service reputation into their recommender system. To solve the cold starting problem in the similarity calculation, they used an item called 'virtual friend'. Firstly, in the sentiment analysis, they built three sentiment dictionaries (SD, SDD and ND) using HowNet sentiment lexicon to calculate the sentiment polarity of the item review. Secondly, they calculated the service reputation factor to use it in the matrix factorization; they used the collaborative filtering algorithm to do this process. They used Yelp dataset in their experiments with eight categories: Active Life, Beauty & Spa, Home Service, Hotel & Travel, Nightlife, Restaurants, Shopping, and pets. They evaluated their work using these two performance measures: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The obtained results show that service reputation has an indirect impact on the rating process, where they achieved improvement of the accuracy compared to similar works.

In [42], the authors improved their previous work [41] where they proposed a sentiment based rating prediction system. They attempt to use these three factors in the recommendation process: the product reputation, sentiments and sentimental interpersonal influence, where they do not rely only on the sentiment of the reviews and that will help improve the rating prediction. They used Latent

Dirichlet Allocation (LDA) [43] to extract the features from the textual reviews. Before that, they pre-processed the data such as removing stop words. After extracting the products' features and the three factors in the consideration, they modeled the matrix factorization. In the experiments, they used around 60,000 friend circles and the items rating by them. Also, the content of dataset is categorized under eight categories: Active Life, Beauty & Spa, Home Service, Hotel & Travel, Nightlife, Restaurants, Shopping, and pets with 300,847 ratings, 96,974 items (with comments at least one comment is posted) and 28,629 users. They evaluated their work by comparing the proposed work with some of the other works, for each category they calculated the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Their work gives significant improvement using this combination of these three factors in their recommender system.

In [44], they proposed an improved recommender system that analyzes the users' ratings to improve the recommendations that target the users and their specific needs. The used dataset was collected from Last.fm. Their evaluation metrics are: accuracy, precision, recall and novelty, where there are some papers that proposed several methods to measure the degree of the novelty such as [45][46][47]. Firstly, in their work they attempt to reduce the dimension of the user-item matrix using the singular value decomposition. Then, they do the clustering phase.

After that, the algorithm is used raise the system experience. Their result outperforms the other methods that used the typical matrix factorization.

In [48], they proposed the Explicit Factor Model (EFM) that helps in generating the recommendation. As a first step, they used the aspect approach to extract the aspect of the reviews. After that, the recommendation system will recommend or not recommend the products to the users with a short explanation for that. They did their result under two conditions: the offline experiment and online experiment. And their evaluation is categorized under two sides: the rating prediction and the top-k recommendation. For the offline experiment, the obtained result outperformed the baseline algorithms. But in the online experiment, the result is varying due to the behavior of the purchaser.

3.1 Analysis of Related works on Reviews Rating Prediction

Table 2 present the differesnt techniques used by authors of all previous works. These techniques can be divided into three main groups:

- Machine learning techniques
- Rule-Based Techniques.
- Lexicon-based Techniques

Ref.	Technique
[27] 2016	VSM + lexicon dictionary
[28] 2016	SVM + lexicon based
[29] 2014	Multi-machine learning algorithms (NB, Ridor, FT, LMT) and lexicon based
[30] 2014	SVM
[32] 2012	SVM, NB
[33] 2015	Lexicon-based
[34] 2014	Lexicon-based
[37] 2016	Arabic sentiment ontology
[38] 2014	Rule-based
[39] 2015	Neural network
[40] 2010	Constrained ridge regression algorithm
[41] 2015	Lexicon-based
[42] 2016	Lexicon-based
[44] 2014	Clustering

Table 2. Analysis of techniques used in Related Works Reviews Rating Prediction Systems

Ref.	Dataset
[27] 2016	ACLIMDB (available online)Vol: 25,000 reviews with labels
[28] 2016	Student reviews collected from:guidetoonlineschools.comVol: 900 reviews
[29] 2014	MPQA (available online)Vol: 9700 sentences
[30] 2014	Same corpus used by Abdul-Mageed et al. (2011)Vol: 400 documents
[32] 2012	Arabic tweets collected from TwitterVol: 1000 Tweets (500 positive and 500 negative)
[33] 2015	Arabic tweets collected from TwitterVol: 4400 Tweets
[34] 2014	DAR: 2798 chat collected from MaktoobTGRD: 3015 tweetsTHR: 3008 sentences collected from WikipediaMONT: 3097 sentences collected from different forums
[37] 2016	Arabic tweets collected from TwitterVol: 1000 Tweets
[38] 2014	- For explicit aspect extraction, they used the dataset that provided by: Hu and Liu and the Semeval 2014 dataset - For implicit aspect extraction, they used the dataset that provided by: Cruz-Garcia et al
[39] 2015	IMDB and two Yelp datasetsVol of:IMDB: 84919 reviewsYelp 2014: 231163 reviewsYelp 2013: 78966 reviews
[40] 2010	Amazon reviews with three domains: books, movies and musicVol: 24000 reviews
[41] 2015	Yelp reviews with 8 categoriesVol: 300847 ratings
[42] 2016	Yelp reviews with 8 categoriesVol: 300847 ratings
[44] 2014	Collected from the webVol: 5000 users with 739817 songs

Table 3. Analysis of Dataset used on Reviews Rating Prediction

Ref.	Accuracy
[27] 2016	Precision, PPV, NPV accuracy.Accuracy for: With stemming = 76.84%, Without stemming= 73.83%
[28] 2016	Precision, recall, F1.F1 for: SVM without subjectivity detection = 78%, SVM with subjectivity detection = 84%
[29] 2014	Precision, recall, F1.F1 for: NB= 70.55%, Ridor= 73.82%, FT= 76.1%LMT= 76%
[30] 2014	Precision, recall, F1.F1 for: Subjective = 72.3%, Positive = 61.6%, Negative = 67.3%, (avg F1= 64.5%)
[32] 2012	Precision, recall, F1, accuracy.F1 for: With stop words: SVM= 72.1%, NB= 65.4%, Without stop words: SVM= 72.5%, NB = 64.6%
[33] 2015	Precision, recall, error rate, accuracy.Precision=70%Recall= 46%
[34] 2014	Accuracy, F1.Accuracy for(with lemmatization):DAR= 70.16%, TGRD= 62.10%, THR= 63.37%, MONT= 72.27%
[37] 2016	Precision, recall.Precision= 75%, Recall= 72%
[38] 2014	Precision, recall.Precision – Recall of: Dataset1: 89.25% - 91.25%, Dataset2: 90.15% - 92.25%, Dataset3: 92.25% - 94.15%, Dataset4: 82.15% - 86.15%, Dataset5: 93.25% - 93.32%
[39] 2015	Accuracy, MAE, RMSE.Accuracy of:Dataset1: 43.5%, Dataset2: 60.8%, Dataset3: 59.6%
[40] 2010	MSE.Dataset1: 0.884, Dataset2: 0.928, Dataset3: 0.627
[41] 2015	MAE, RMSE.With 8 datasets used, the average of :MAE= 1.007RMSE= 1.272
[42] 2016	Precision, MAE, RMSE.Precision of: Dataset1: 72.7%, Dataset2: 73.5%, Dataset3: 87.1%
[44] 2014	Precision, recall.Precision = [0.0085 – 0.0123], Recall= [0.0011 – 0.0016]

Table 4. Analysis of Accuracy of Reviews Rating Prediction Systems

Ref.	Advantages
[27] 2016	The results show SA has a positive effect in the rating prediction
[28] 2016	Using subjectivity detection as pre-step helps in improving the classification performance
[29] 2014	Their lexicon SentiRDI can be used as Arabic lexicon for SA
[30] 2014	Their large scale Arabic lexicon is publically available
[32] 2012	In their research, SVM has better performance in SA than NB
[33] 2015	The result shows, using lexicon in sentiment detection is helpful
[34] 2014	Better to do lemmatization
[37] 2016	Using ontology in sentiment analysis improve the classification performance
[38] 2014	Their aspect extraction framework can extract implicit and explicit aspects using rule-based method
[39] 2015	Neural network can achieve high performance in field of document level sentiment
[40] 2010	Bag of opinion method is suitable for the sparse datasets
[41] 2015	They used additional factors (rather than the traditional ones) in their recommender system to enhance the performance.
[42] 2016	Improvement of [41], with one more factor to achieve better performance of the recommendation
[44] 2014	Clustering items helps in improving the recommendation process performance

Table 5. Analysis of advantages of Reviews Rating Prediction Systems

Table 3 present the Datasets used by different authors in their works. We can define two groups of datasets:

- Public Academic Datasets
- Professional Datasets.

Table 4 present the Accuracy in terms of Precision, recall and F1. These measures are taken for different datasets and different techniques.

Table 5 presents the different advantages and specifications of related works:

3.2 Results Discussion of Related works on Reviews Rating Prediction

From Tables 2, 3, 4, 5, we can conclude some important notes:

- The preprocessing is considered as an important step in the sentiment analysis.
- The most used machine learning algorithm is SVM since it works well with the binary classification problem.
- In document level sentiment analysis, neural networks can work well.
- The datasets used in the afore mentioned studies are different and vary in their characteristics that make the comparison in term of dataset very difficult.
- The most used evaluation metrics are accuracy, F1 measure, precision and recall.

- The existing studies that combine recommender systems with sentiment analysis are in general for English language, so this point motivated us to opt for Arabic language in our study.

4. Proposed System

In general, our proposed system aims to combine the recommender systems with sentiment analysis systems in one hybrid system. It intends to predict the ratings using the polarity of the Arabic reviews. This section will present the system major phases and components.

4.1 System Description

Our objective in this project is to have a recommender system that considers the sentiment of the customer reviews. In simple words, in the star prediction process the sentiment of the reviews will be considered and counted where as a valuable input information for our system; where in the traditional recommender systems, usual factors and information such as user information are considered. The textual review is an important element in the composition of the customer's viewpoint. For this reason, we aim to use such information in our recommender system. Therefore, our system will use an Arabic reviews dataset as input, and use the output of the sentiment analysis process as input for the recommendation process.

4.2 System Architecture

The system, from its general view, consists of two following main components: sentiment analyzer and the recommender system (star predictor). These two components are shown in Figure 6 (high abstract block diagram)

No.	Component	Function
1	Text preprocessing	Preprocess the textual reviews e.g. clean the text from the noise, remove stop words, stemming, etc. – Figure 9.
2	Sentiment analyzer	After preprocessing the text, here is the analysis of the reviews sentimentally, using machine learning supervised algorithm (SVM) – Figure 8
3	Star rating predictor	Predict the star rating of the product based on multi-input (direct rating and indirect rating from the text) – Figure 10.

Table 6. Component of the system

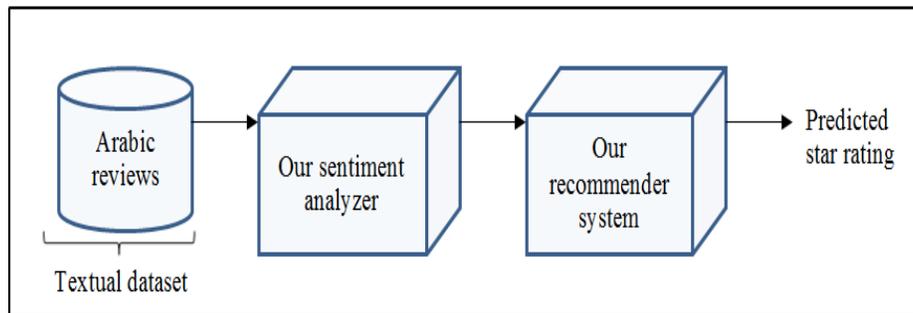


Figure 6. System block diagram

that represents only the phases of the system. Where the input of the system is the textual Arabic reviews dataset and the output of the system is the star rating. For the recommender phase, the input will be the same output from the previous phase i.e. the review polarity that will be calculated by the sentiment analyzer; that will be used as one of the input to the RS factorization matrix.

4.3 System Components

Our proposed system components can be listed as follows:

- Text preprocessing.
- Sentiment analyzer.
- Star rating predictor.

In Table 6, there is a brief description of each component function.

4.4 Illustrative Example

In Figure 7, we have an Arabic textual review example, with the star rating from the user. Here, the user's direct rating Ur is equal to 1 star rating. By looking at the review, we can find some important keywords (features), such as: *مب دقيق، خفيفه، مريحه،...* etc. After analyzing them sentimentally, we will have Rp (+1 or -1) to be used in the star prediction phase to get the total rating value. After the calculations, it will be the input to the factorization matrix and then the star rating of this product can be later predicted using model-based collaborative filtering.

4.5 System Implementation Phases

4.5.1 Datasets and lexicons

There are several Arabic datasets and lexica that are publicly available for research purposes.

Here are some Arabic reviews datasets and corpora such as OCA which includes 500 reviews that were collected from websites and blogs and it involves movies and books reviews [49]. Also, HAAD which consists of 2389 reviews from several websites and it is about book reviews [50].

Both, OCA and HAAD are MSA dialect. LABR is an Arabic reviews dataset that has over 63000 reviews about books that were collected from www.goodreads.com with general dialect [51].

On the other hand, there are several Arabic lexica such as SLSA which consists of 35000 words in MSA and it was constructed using AraMorph and SentWordNet [52]. Also, AOR lexicon which combines MSA and Saudi dialect since it has 1000 words from Arabic MPQA (MSA) and 2690 Saudi dialectal words from Twitter [53]. Another lexicon SentiRDI, where it consists of 3256 positive words, 4169 negative and 10839 neutral words with MSA [29] and other lexica.

Figure 8 shows a sample of an Arabic review structure. The review is composed of text review section and star rating section.

This later represents the number of stars given by a customer buying the product. In our system, the textual re-

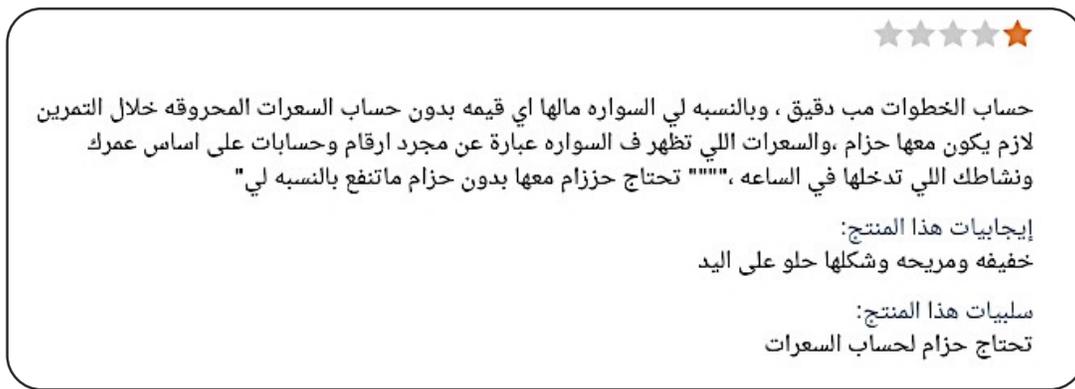


Figure 7. An example of Arabic text review

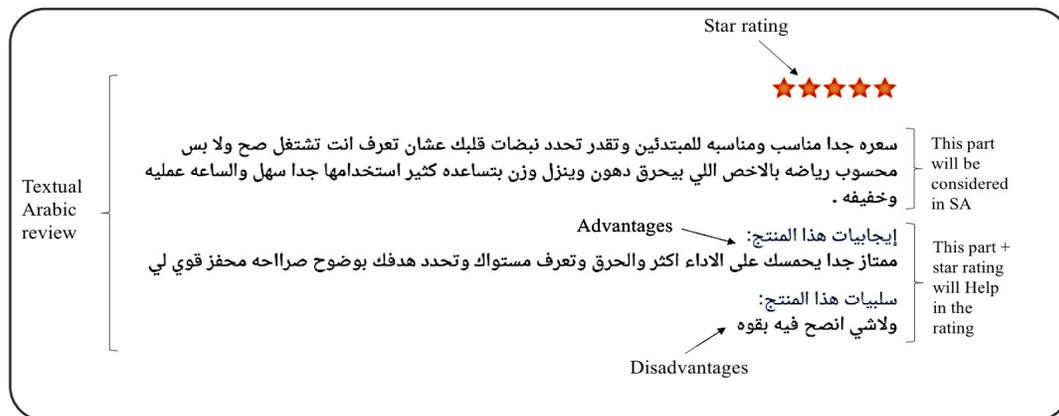


Figure 8. Arabic review sample

view will be considered for the SA phases (as an indirect rating) as shown in Figure 8; the advantages, disadvantages and star rating help in the star rating (will be used as features input for the direct rating).

4.5.2 Phase of Sentiment Analysis

In Figure 9, we give more details about the sentiment analysis and NLP phases. But before using this textual reviews in this phase, we must preprocess the data.

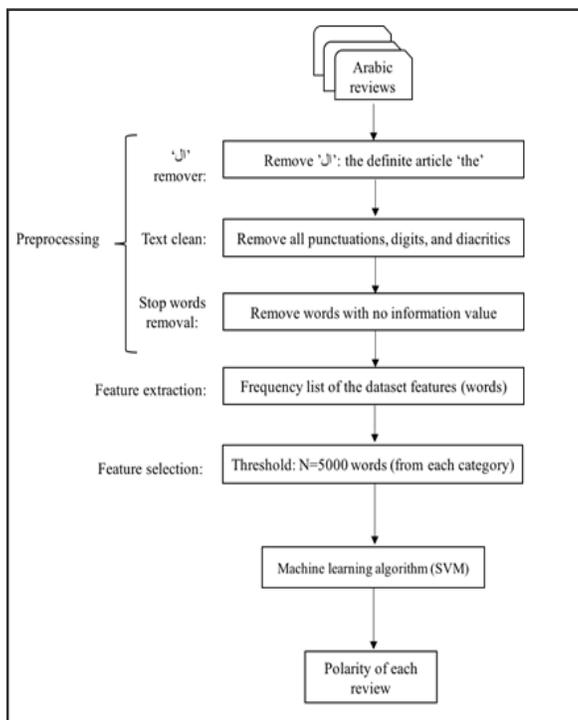


Figure 9. Sentiment analysis phase

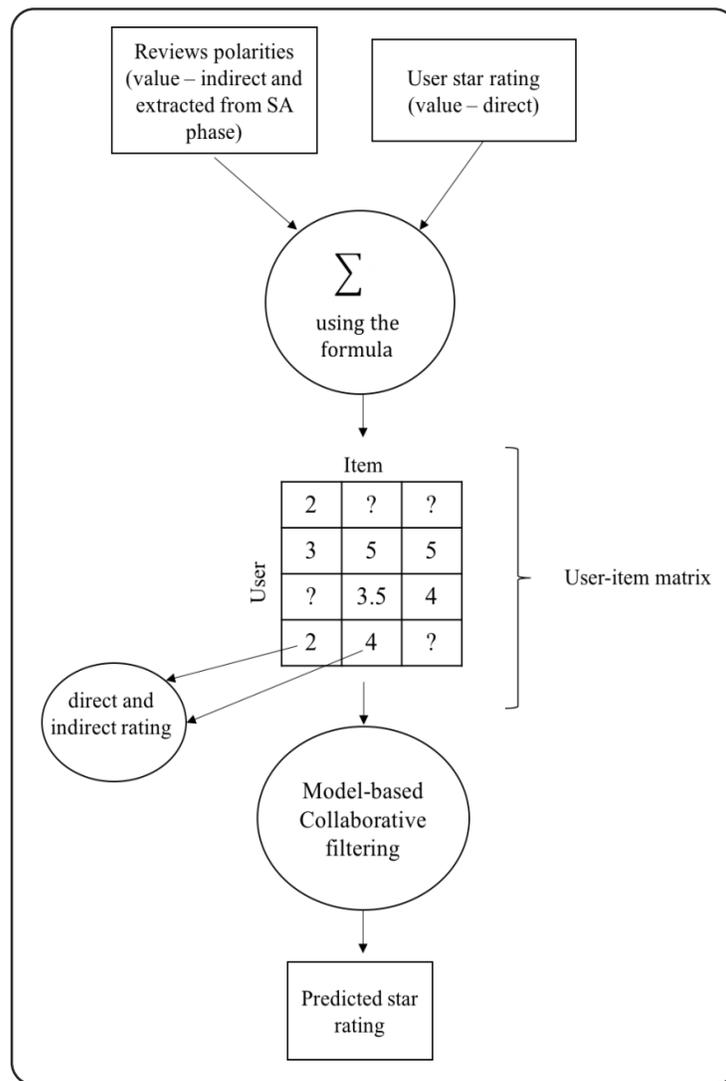


Figure 10. Star prediction phase

4.5.3 Phase of Star Prediction

Rather than using only the default criteria in the prediction matrix such as the item and user information, we attempt to add the polarity of the reviews to it. The prediction process depends on the factorization matrix where some values are existing and others are missing. We aim to predict these missing values as shown in Figure 10.

Figure 10 shows the whole processes for this phase. The input for the prediction matrix will be the two following values: the user star rating (direct rating) and the polarity from the text reviews (indirect rating, which comes from the previous phase as +1 or -1). So, we used these two kinds of rating as combined input. The direct rating came from the user, and the indirect rating from the textual reviews.

4.6 Evaluation Metrics

There are several evaluation metrics for different kinds of applications such as: Information Retrieval (IR), information extraction, SA, RS. Evaluation is an important phase. For that aim, we will use different kinds of metrics like:

accuracy, precision, recall, F1 and RMSE which will be listed below with more details.

5. Conclusion and Future Work

In this study, we reviewed different methods and techniques in the field of sentiment analysis (SA) and recommender system (RS). Most of the recent studies in SA rely on using machine learning or/and lexicon based approach to enhance the sentiment analysis accuracy. On the other hand, we presented the methods of recommender systems such as collaborative filtering (memory-based and model based), content-based and knowledge-based. Also, challenges that faced by both SA and RS has been reviewed. The main contribution of this project is to combine these two fields and develop one hybrid solution that has the ability to predict rating of products by considering the sentiment of these products reviews.

This study could be the base for many other future works by using larger dataset with larger number users and items (such as movies, products or books). More experiments

can be conducted by tuning parameters of feature selection, testing the behavior of the system and comparing the obtained results.

References

- [1] The Role of Customer Product Reviews - eMarketer, Emarketer.com, 2017. [Online]. Available: <https://www.emarketer.com/Article/Role-of-Customer-Product-Reviews/1008019>. [Accessed: 06- Apr- 2017].
- [2] Gesenhues, A. (2017). Survey: 90% Of Customers Say Buying Decisions Are Influenced By Online Reviews, Marketing Land, 2017. [Online]. Available: <http://marketingland.com/survey-customers-more-frustrated-by-how-long-it-takes-to-resolve-a-customer-service-issue-than-the-resolution-38756>. [Accessed: 06- Apr- 2017].
- [3] Web Users Put More Stock in Consumer Reviews - eMarketer, Emarketer.com, 2017. [Online]. Available: <https://www.emarketer.com/Article/Web-Users-Put-More-Stock-Consumer-Reviews/1012929>. [Accessed: 06- Apr- 2017].
- [4] Internet Users Rely on Reviews When Deciding Which Products to Purchase - eMarketer, Emarketer.com, 2017. [Online]. Available: <https://www.emarketer.com/Article/Internet-Users-Rely-on-Reviews-Deciding-Which-Products-Purchase/1014465>. [Accessed: 06- Apr- 2017].
- [5] Local Consumer Review Survey 2016 | The Impact Of Online Reviews, BrightLocal, 2017. [Online]. Available: <https://www.brightlocal.com/learn/local-consumer-review-survey/>. [Accessed: 06- Apr- 2017].
- [6] Liu, Bing. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1) 1–167.
- [7] Liu, Bing. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.
- [8] Schouten, K., Frasincar, F. (2016). Survey on Aspect-Level Sentiment Analysis, *IEEE Transactions on Knowledge and Data Engineering*, 28 (3) 813-830.
- [9] Che, W., Zhao, Y., Guo, H., Su, Z., Liu, T. (2015). Sentence Compression for Aspect-Based Sentiment Analysis, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23 (12) 2111-2124.
- [10] Sarawgi, K., Pathak, V. (2017). Opinion Mining: Aspect Level Sentiment Analysis using SentiWordNet and Amazon Web Services, *International Journal of Computer Applications*, 158 (6) 31-36.
- [11] Alhojely, S. (2016). "Different Applications and Techniques for Sentiment Analysis, *International Journal of Computer Applications*, 154 (5) 24-28.
- [12] Altwairesh, N. S. (2016). Sentiment Analysis of Twitter: A Study on the Saudi Community," Ph.D dissertation, KSU Univ., Riyadh, 2016.
- [13] Ayyoub, M., Essa, S., Alsmadi, I. (2015). Lexicon-based sentiment analysis of Arabic tweets, *International Journal of Social Network Mining*, 2 (2) 101, 2015
- [14] AL-Rubaiee, H., Qiu, R., Li, D. (2016). The Importance of Neutral Class in Sentiment Analysis of Arabic Tweets, *International Journal of Computer Science and Information Technology*, 8 (2) 17-31.
- [15] Ben Salamah, J., Elkhilfi, A. (2014). Microblogging Opinion Mining Approach for Kuwaiti Dialect, *In: International Conference on Computing Technology and Information Management*, 2014.
- [16] Al-Kabi, M., Alsmadi, I., Gigieh, A., Wahsheh, H. and Haidar, M. (2014). Opinion Mining and Analysis for Arabic Language, *International Journal of Advanced Computer Science and Applications*, 5 (5).
- [17] Aggarwal, C. (2016). Recommender systems, 1st ed.
- [18] Kembellec, G., Chartron, G., Saleh, I. (2015). Recommender Systems, 1st ed. John Wiley & Sons.
- [19] Tugaonkar, P., Chitre, V. (2016). Survey on recent methodologies used for recommender system, *In: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016.
- [20] Isinkaye, F., Folajimi, Y., Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation, *Egyptian Informatics Journal*, 16 (3) 261-273.
- [21] Yang, Z., Wu, B., Zheng, K., Wang, X., Lei, L. (2016). A Survey of Collaborative Filtering-Based Recommender Systems for Mobile Internet Applications, *IEEE Access*, vol. 4, p. 3273-3287.
- [22] Ashley-Dejo, E., Ngwira, S., Zuva, T. (2015). A survey of Context-aware Recommender System and services, *In: 2015 International Conference on Computing, Communication and Security (ICCCS)*, 2015.
- [23] Yadav, P. (2016). Recommender System: Revolution in E-Commerce, *International Journal Of Engineering And Computer Science*, 2016.
- [24] Ambulkar, H., and Pathan, A. (2015). Recommender System Challenges and Methodologies in Social Network: Survey, *International Journal of Science and Research (IJSR)*, 4 (11) 286-289.
- [25] Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., Duval, E. (2012). Context-Aware Recommender Systems for Learning: A Survey and Future Challenges, *IEEE Transactions on Learning Technologies*, 5 (4) 318-335.
- [26] Lerato, M., Esan, O., Ebunoluwa, A., Ngwira, S., Zuva, T. (2015). A survey of recommender system feedback techniques, comparison and evaluation metrics, *In: 2015 International Conference on Computing, Communication and Security (ICCCS)*, 2015.
- [27] Alshari, E., Azman, A., Mustapha, N., Doraisamy, S., Alksher, M. (2016). Prediction of rating from comments based on information retrieval and

- sentiment analysis, *In: 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP)*.
- [28] Omar, N., Eesee, A. (2016). A Hybrid Method for Arabic Educational Sentiment Analysis, *Journal of Applied Sciences*, 16 (5) 216-222.
- [29] Mobarz, H., Rashown, M., Farag, I. (2014). Using Automated Lexical Resources in Arabic Sentence Subjectivity, *International Journal of Artificial Intelligence & Applications*, 5 (6) 01-14.
- [30] Badaro, G., Baly, R., Hajj, H., Habash, N. El-Hajj, W. (2014). A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining, *In: Proceedings of the EMNLP Workshop on Arabic Natural Language Processing (ANLP)*, 2014, p.165-173.
- [31] Almuqren, L., Cristea, A. (2016). Framework for Sentiment Analysis of Arabic Text, W. *In: Proceedings of the 27th ACM Conference on Hypertext and Social Media - HT '16*, 2016.
- [32] Shoukry, A., Rafea, A. (2012). Sentence-level Arabic sentiment analysis, *In: 2012 International Conference on Collaboration Technologies and Systems (CTS)*, 2012.
- [33] Duwairi, R., Ahmed, N., Al-Rifai, S. (2015). Detecting sentiment embedded in Arabic social media – A lexicon-based approach, *Journal of Intelligent & Fuzzy Systems*, . 29 (1) 107-117.
- [34] Abdul-Mageed, M., Diab, M., Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media, *Computer Speech & Language*, 28 (1) 20-37.
- [35] Younis, M.G.E. (2015). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study, *International Journal of Computer Applications*, 112 (5).
- [36] Al-Twairsh, N., Al-Khalifa, H., Al-Salman, A. (2015). Towards Analyzing Saudi Tweets, 2015 First International Conference on Arabic Computational Linguistics (ACLing), 2015.
- [37] Tartir, S., Abdul-Nabi, I. (2016). Semantic Sentiment Analysis in Arabic Social Media, *Journal of King Saud University - Computer and Information Sciences*, 2016.
- [38] Poria, S., Cambria, E., Ku, L., Gui, C., Gelbukh, A. (2014). A Rule-Based Approach to Aspect Extraction from Product Reviews, *In: Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP)*, 2014.
- [39] Tang, D., Bing, Q., Liu, T. (2015). Learning semantic representations of users and products for document level sentiment classification, *In: Proc. 53th Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China, 2015, p. 1014–1023.
- [40] Qu, L., Ifrim, G., Weikum, G. (2010). The bag-of-opinions method for review rating prediction from sparse text patterns, *In: Proc. 23rd International Conference on Computational Linguistics*, 2010, p. 913–921.
- [41] Lei, X., Qian, X. (2015). Rating prediction via exploring service reputation, *In: 17th International Workshop on Multimedia Signal Processing (MMSp)*, 2015.
- [42] Lei, X., Qian, X., Zhao, G. (2016). “Rating Prediction Based on Social Sentiment From Textual Reviews, *IEEE Transactions on Multimedia*, 18 (9) 1910-1921.
- [43] Blei, D.M., Ng, A.Y., Jordan, M. I. (2003). Latent Dirichlet Allocation, *Journal of Machine Learning Research* 3. 2003, p. 993-1022.
- [44] Lee, K., Lee, K. (2014). Using Dynamically Promoted Experts for Music Recommendation, *IEEE Transactions on Multimedia*, 16 (5) 1201-1210.
- [45] Akiyama, T., Obara, K., Tanizaki, M. (2010). Proposal and evaluation of serendipitous recommendation method using general unexpectedness, *In: Workshop on the Practical Use of Recommender Systems, Algorithms and Technologies (PRSAT 2010)*, 2010, p. 3.
- [46] Murakami, T., Mori, K., Orihara, R. (2008). Metrics for evaluating the serendipity of recommendation lists, *New Frontiers in Artificial Intelligence*, p. 40–46, 2008.
- [47] Zhou, T., Kuscsik, Z., Liu, J., Medo, M., Wakeling, J., Zhang, Y. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems, *In: Proceedings of the National Academy of Sciences*, 107 (10) 4511–4515.
- [48] Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., and Ma, S. (2014). Explicit factor models for explainable recommendation based on phrase-level sentiment analysis, *In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '14*, 2014.
- [49] Rushdi-Saleh, M., Martín-Valdivia, M., Ureña-López, L., Perea-Ortega, J. (2011). OCA: Opinion corpus for Arabic, *Journal of the American Society for Information Science and Technology*, 62 (10).
- [50] Al-Smadi, M., Qawasmeh, O., Talafha, B., Quwaider, M. (2015). Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis, *In: 2015 3rd International Conference on Future Internet of Things and Cloud*, 2015.
- [51] Aly, M., Atiya, A. (2013). LABR: A Large Scale Arabic Book Reviews Dataset, *ACL*, 2013.
- [52] Eskander, R., Rambow, O. (2015). SLSA: A Sentiment Lexicon for Standard Arabic, *In: Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 2015.
- [53] Alhumoud, S., Albuhairi, T., Alohaideb, W. (2015). Hybrid sentiment analyser for Arabic tweets using R, *In: Knowledge Engineering and Knowledge Management (IC3K)*, Lisbon, Portugal, 2015.