**A Context-aware Recommendation Framework with Explicit Sentiment Analysis**

## A SEMINAR REPORT

*Submitted by*

## Mable Biju

**MBT16CS068**

*In partial fulfilment of the requirements for the award of*

*the degree of*

## BACHELOR OF TECHNOLOGY

### in

**COMPUTER SCIENCE AND ENGINEERING**

**Of the A.P. J. Abdul Kalam Technological University**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING MAR BASELIOS COLLEGE OF ENGINEERING AND TECHNOLOGY MAR IVANIOS VIDYANAGAR, NALANCHIRA THIRUVANANTHAPURAM – 695 015**

**November 2019**

**Mar Baselios College of Engineering & Technology**

# Mar Ivanious Vidyanagar, Nalanchira,

**Thiruvananthapuram - 695 015**

## (Affiliated to A. P. J. Abdul Kalam Technological University)

**Department of Computer Science and Engineering**



**CERTIFICATE**

This is to certify that the Seminar Report titled ‘*A Context-aware Recommendation Framework with Explicit Sentiment Analysis’* is a bonafide record of seminar presented by Mable Biju (MBT16CS068).

|  |  |  |
| --- | --- | --- |
| **Mrs. Jesna Mohan** | **Mrs. Shini Renjith** | **Dr. Tessy Mathew** |
| Seminar Co-ordinator | Seminar Guide | Head of the Department |

### Place: Thiruvananthapuram Date: / /2019

**ACKNOWLEDGEMENT**

First and foremost, I thank God Almighty for His divine grace and blessings in making all these possible. May He continue to lead us in the years. It is to render my heartfelt thanks and gratitude to our principal**, Dr. T. M. George** for providing the opportunity to do this seminar during the 7th semester (2019) of my B. Tech degree course.

I am deeply thankful for our Head of the Department**, Dr. Tessy Mathew** for her support and encouragement. I would like to express my sincere gratitude to my seminar guide Asst. Professor, **Mrs. Shini Renjith**, and project coordinator, Asst. Professor, **Mrs. Jesna Mohan,** Department Computer Science and Engineering for their motivation, assistance and help for the seminar. I also thank all the staff members of the Computer Science Department for providing their assistance and support. Last, but not the least, I thank all my friends and family for their valuable feedback from time to time as well their help and encouragement.

## Mable Biju

**CONTENTS**

1. [Introduction 7](#_TOC_250008)
2. [Motivation 9](#_TOC_250007)
3. [System Framework 1](#_TOC_250006)0
   1. Review Pre-processing Module 11

3.A.1. Word Stemming 12

3.A.2. Tokenization 12

3.A.3. Stop Word Removal 12

3.A.4. Word Spelling 12

* 1. Review Analysis Module 13

3.B.1. PoS Tagging 13

3.B.2. Feature Extraction 13

* 1. Polarity Detection Module 14

3.C.1. Naïve Bayes Model 15

3.C.2. SVM Model 15

3.C.3. Aggregated Sentiment Score 15

* 1. Recommendation Module 16

3.D.1. Popularity Ranking 16

1. [Comparison 17](#_TOC_250001)
2. [Conclusion 19](#_TOC_250000)
3. References 20

**List of Figures**

FIGURE 2.1. Example of biasness in user rating 9

FIGURE 3.1. Recommendation system framework 10

FIGURE 3.2. PoS Tagging done on a sentence 13

FIGURE 3.3. Polarity Detection performed using machine learning 14

FIGURE 4.1. Model Comparison 17

**Abstract**

*The recommendation systems usually match users’ preferences based on the star ratings provided by the users for various products. However, simply relying on users’ ratings about an item can produce biased opinions, as a user’s textual feedback may differ from the item rating provided by the user.*

*SocialRec, a hybrid context-aware recommendation framework that utilizes a rating inference approach to incorporate users’ textual reviews into traditional collaborative filtering methods for personalized recommendations of various items.*

*The motivation behind this is to bridge the gap between recommendations based on numeric ratings, and those of users’ opinions entered as free text to reach an unbiased rating about an item/venue.*

*Here text-mining algorithms are applied on a large-scale user-item feedback dataset to compute the sentiment scores. Moreover, the proposed framework consists of 4 preprocessing modules which makes it overcome some of the conventional issues of recommenders.*

*Rigorous evaluations of SocialRec (on large-scale datasets) demonstrate high accuracy, especially in comparison with previous related framework.*

# 1. INTRODUCTION

Recommender system  is a subclass of information filtering system that seeks to give a general recommendation based on the rating and the textual reviews combined.

There are 2 types of recommender system:

1.Content-based: The basic idea of content-based filtering is that if you like an item you will also like a ‘similar’ item.  The content of each item is stored in the database, this is further .

2.Collaborative-Filtering based (CF):  It is based on the assumption that people who agreed in the past will agree in the future. This filtering method is usually based on collecting and analyzing information on user’s behaviors, their activities or

preferences and predicting what they will like based on the similarity with other users.

There are also several other categories of recommenders like knowledge-based, demographic-based and hybrid.

Some common issues regarding recommenders are:

1. Cold start- Lack of information regarding new items.
2. Data sparsity- Large amount of items, but less user rating.
3. Scalability- Growth of number of users and items are not proportional.

The best way to tackle this issue is by using a hybrid recommender system. This system combines two or more standard recommendation systems to benefit from their complementary advantages.

One of the most popular example of a hybrid recommender system is Netflix.

It makes recommendations by comparing the watching and searching habits of similar users (i.e., collaborative filtering). It offers movies that share characteristics with films that a user has rated highly (content-based filtering).

Some recent technologies that are used in this proposed system are:

1.CONTEXT AWARENESS

It is the ability of a system or system components to gather information about its environment at any given time and adapt behaviors accordingly. It uses software and hardware to automatically collect and analyze data to guide responses.

In short, context awareness is the fact of being aware of the location or surroundings and making appropriate decisions.

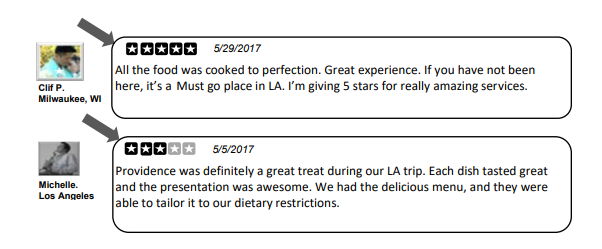
2. SENTIMENT ANALYSIS

It identifies and extracts subjective information in source material. It helps a business to understand the social sentiment of their brand, product or service while monitoring online conversations. However, analysis of social media streams is usually restricted to just basic sentiment analysis and count based metrics.

# MOTIVATION

The main motivation behind such framework is the biasness caused by user star rating and the actually opinion. Such a mismatch can lead to the degradation of the actual recommendation and the entire recommender system.

From the figure given below, the two reviewers have given their review for a particular restaurant.



**FIGURE 2.1** Example of biasness in user rating

From the above figure it is clear that even though both users gave positive feedback, the star rating was different. This leads to a biasness in the actual recommendation system.

Many conventional Collaborative filtering systems use the 5-star rating system to explicit generate a recommendation. Moreover, users also prefer a text free rating. However, these conditions increase biasness and affect the essence of the review.

In this way, it is critical to cross over any barrier between suggestions dependent on numeric evaluations, and those of users' textual review entered as free content to arrive at an unprejudiced rating about an item or venue.

# 3. SYSTEM FRAMEWORK

The system consists of a polarity detection module which is used for sentimental analysis and a record is used to store this score. Along with this a separate record is also maintained to keep the reviews and ratings of each user and venue.

This entire framework is divided into four phases as follows:

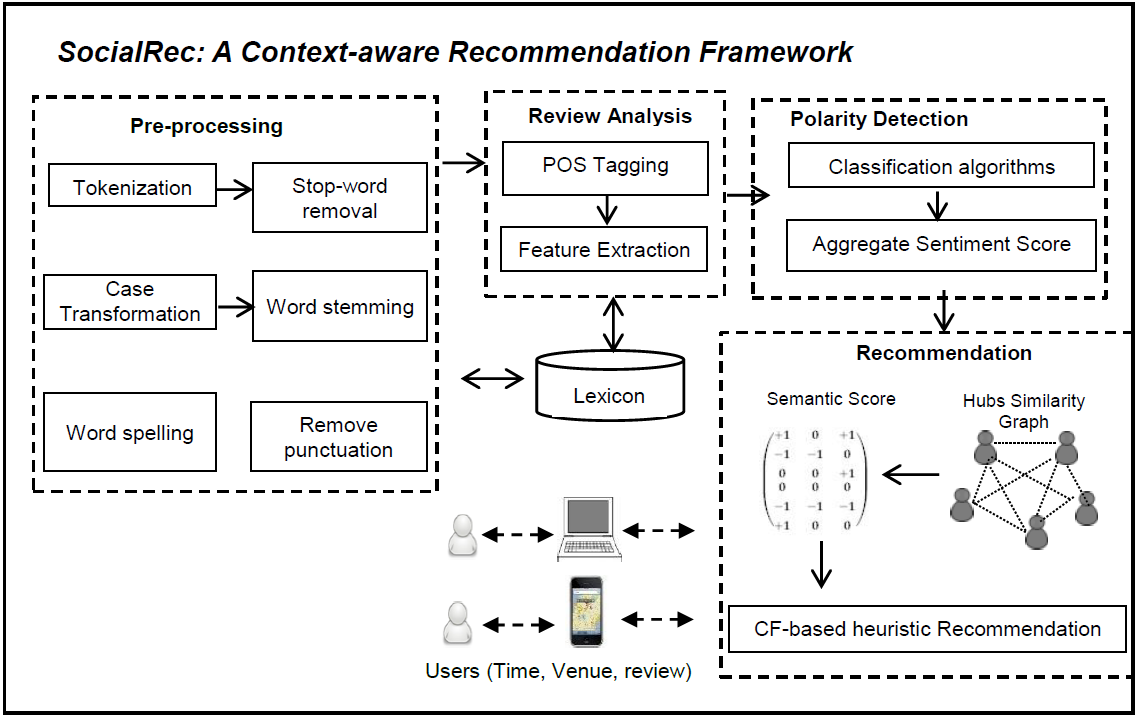
3.A. Review pre-processing module

3.B. Review analysis module

3.C. Polarity detection module

3.D. Recommendation module

The figure given below shows the entire architecture of the recommender system.



**FIGURE 3.1** Recommendation system framework

**3.A. Review Pre-processing Module**

This module ensures that the unstructured review is processed and converted into a proper format before performing the machine learning algorithms for the actual recommendation system. This phases removes all the noisy and unclean data from the first stage itself.

The following are the various stages of pre-processing

**3.A.1. Word stemming**

It is the process of removing all the tenses, suffixes and prefixes from a word. In such a way the root word can be obtained. This process reduces the time taken for the pre-processing phase.

**3.A.2. Tokenization**

This step is done to divide the words of each sentences into different categories like numerical, punctuations, words, etc. The entire process of review pre-processing would be much easier after the tokenization.

**3.A.3. Stop word removal**

Stop words are those words that occur repeatedly and does not provide any significant syntactic difference to the sentence. Some of the stop words are “the”, “am”, “a”, “an”, etc. As a result, it is necessary to remove stop words from the text corpus.

**3.A.4. Word spelling**

Words that are incorrectly spelt are a main hindrance to the review analysis phase. So it is indeed necessary to check for the correct spelling in the textual review. PyEnchant, is a free python library that be used for spell check. It has a predefined dictionary from where it will replace the misspelt word with the most probable correct word.

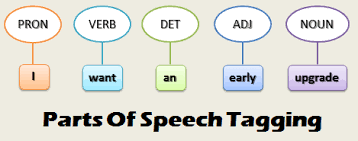
Some additional steps such as case transformation and punctuation removal are also performed in this module. The main aim of review pre-processor is to make the raw textual input suitable for review analysis phase.

**3.B. Review Analysis Module**

The review analysis phase examines the linguistic features of review so that from the review a particular opinion can be determined. Two significantly received assignments for review analysis examination are POS labeling and include extraction. The description depiction of the previously mentioned is exhibited in ensuing content.

**3.B.1. PoS tagging**

This is one of the most important step in this particular framework. It is done using NLTK and each sentence is split after the review is parsed completely. Subsequently, a tag is added to each word in that sentence. This tag will indicate the part of speech that the word denotes in that particular sentence. It can be a noun, verb, adjectives, etc. After the Pos tagging, the review along with the tag is used for feature extraction.



**FIGURE 3.2** PoS tagging done on a sentence

**3.B.2. Feature extraction**

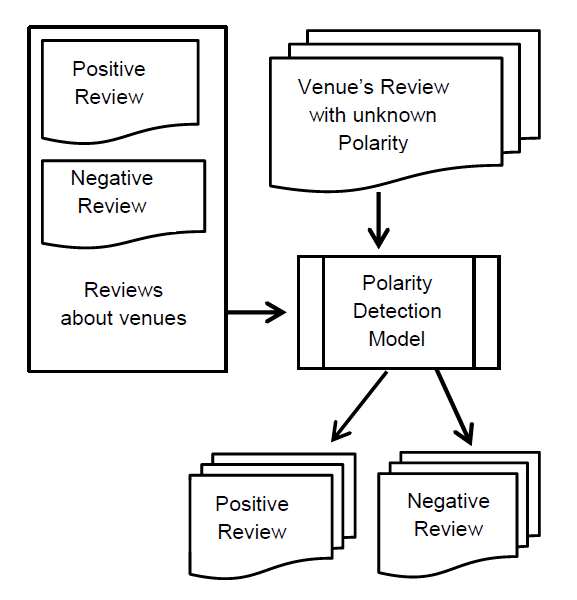
Sentences can be broadly classified into two types, objective and subjective sentences. Feature extraction is the process of identifying these types of sentences. Subjective sentences are those which states hard facts and objective sentences are those which state emotions of feelings. In the context of recommendation system, we only need the subjective type sentences. So after the feature extraction phase the objective sentences are pruned out. SentiWord Net is used for sentiment analysis as it is a lexical resource.

After the review analysis module, the dimension of the entire input is reduced significantly. This is an added benefit to the whole recommendation system compared to the conventional systems.

**3.C. Polarity Detection Module**

At the end of this module, there is a sentence-level classification. The reviews are classified as positive, negative or neutral after this module. There will be fine grained polarity detection since there is sentence-level pruning.

Various supervised learning algorithms can be implemented for the polarity detection module. Naïve Bayes and SVM are used because they have better results in text mining.



**FIGURE 3.3** Polarity Detection performed using machine learning

**3.C.1. Naïve Bayes Model**

There are two types of distribution models for sentiment classification, probability based and deterministic model. Naïve Bayes is a probabilistic model for classification. A probability is given to each review based on the similarity with the output class. The output class with the highest probability will be chosen at the end.

**3.C.2. SVM Model**

It is one of the most popular machine learning algorithm that is used for text based inputs. This method makes use of drawing a hyperplane that separates the various classes of output. Based on this the output is assigned. The distance between the two hyperplanes represents the margin of the decision boundary.

**3.C.3. Aggregated Sentiment Score**

Once the classification is done, each sentence will have a sentiment score like 1,0 and -1 for positive, neutral and negative sentences respectively. Since a review is not just made up of one sentence, it will require an aggregated score for all the sentences in that review. This is done by adding up the sentiment score of each sentence in the review. After pruning through the entire review if the final aggregated score is more zero, then it’s a positive number. If it is less than zero, then it is a negative review. Otherwise it will be a neutral review.

After the polarity detection module, the aggregated score is used for popularity ranking in the recommendation phase. Once the sentences are found to be neutral, it will be extracted out just like objective sentences as they contain no vital information for the rest of the recommendation process.

**3.D. Recommendation Module**

In this module, the input will be the sentiment score from the polarity detection module. As an output, it recommends for an active reviewer the top-N venues. Here N is the number of venues for that particular user. Further HA-based inference model is used to compute numerical ranks of reviewers and venues. The basic idea of the HA-based inference model is to assign ranking to the reviewers and venues based on a mutual reinforcement relationship.

**3.D.1. Popularity Ranking**

A reviewer is assigned a higher rank (or called as expert), if the reviewer has given reviews for about numerous high ranked venues. Likewise, a venue gets a higher score (or called as popular) if the venue is given review by many expert reviewers.

Further, a similarity graph is computed by the recommendation module of the expert reviewers. In order to reduce the online processing time, the reviews with low scores are pruned out in this module. Thereby the dataset becomes more compact.

Heuristic based approach finds out the venues that matches the user’s preferences the best as suggestions. The suggestions that satisfies the user the most will be the ones in the top of the recommendation list that is made after popularity ranking.

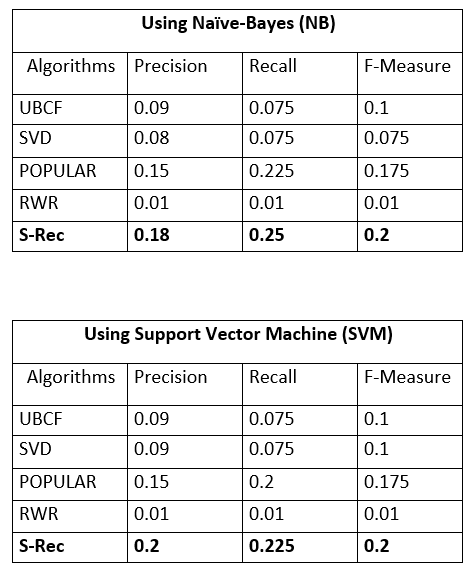
This module is the last one in the framework and this provides the proper recommendation for the user based on context awareness and explicit sentiment analysis.

# COMPARISON

# A comparison was made with this framework and several other existing frameworks like UBCF (User-based Collaborative Filtering), (b) SVD (Singular Value Decomposition), RWR (Random Walk with Restart), and Popular. Several performance metrics such as precision, recall and f-measure was used to evaluate their performance.

# A standard 10-fold cross validation technique was used and 80% of the data was taken for training with just 20% for testing.

# The results can be observed in the following figure.



**FIGURE 4.1** Model Comparison

It is clear that for both Naïve Bayes and SVM based classification, the SocialRec stood out with clear lead margins. Moreover, it also outperforms the other schemes in terms of all the performance metrics.

This model has also addressed some of the conventional problems of recommendation systems. Cold start problem can be removed to some extent by the use HA based approach. Data sparsity can be tackled by using conditional probability as in the case of Naïve Bayes classifier.

Another advantage of this system is that, even at high number of recommendations the performance of the system do not degrade. As this system uses top-k based recommendation, the k value plays a very important job in terms of precision, recall and f-measure. As the value of k increases most of the recommendation system tends to give low results. However, even at *K=20*, the average overall performance of this system remains superior than the existing schemes.

# CONCLUSION

This seminar was about recommendation system that uses context awareness and explicit sentiment analysis. Conventional recommenders do not have a mechanism to detect biasness in terms of the contradicting ratings and textual reviews. However, by using intense data pre-processing and polarity detection for sentiment score we can prune out the unwanted reviews and at the same time bridge the gap between star based rating and textual reviews. This system also outperforms the existing systems in other performance evaluating measures

# 6. REFERENCES

[1] R. Irfan, O. Khalid, M. U. S. Khan, F. Rehman, A. U. R. Khan and R. Nawaz, "SocialRec: A Context-Aware Recommendation Framework With Explicit Sentiment Analysis," in IEEE Access, vol. 7, pp. 116295-116308, 2019.

[2] L. Cui, W. Huang, Q. Yan, F. R. Yu, Z. Wena and N. Lu, "A novel context-aware recommendation algorithm with two-level SVD in social networks," *Future Generation Computer Systems,* vol. 86, p. 1459–1470, 2018.

[3] S. N. Alves-Souza, L. V. L. Filgueiras and L. S. DeSouza, "Sentiment Analysis of Social Network Data for Cold-Start Relief in Recommender Systems," in *WorldCIST*, 2018.

[4] Q. Qi, Z. Chen, J. Liu, C. Hui and Q. Wu, "Using Inferred tag rating to Improve User-based Collaborative Filtering," in Proceedings of the 27th ACM Symposium on Applied Computing, ACM, 2012.

[5] M. Eirinaki, J. Gao, I. Varlamis and K. Tserpes, "Recommender Systems for Large-Scale Social Networks: A review of challenges and solutions," Future Generation Computer Systems, vol. 78, no. 1, pp. 403-418, 2018.