

Feature Level Opinion Mining of Educational Student Feedback Data using Sequential Pattern Mining and Association Rule Mining

Ayesha Rashid

Faculty of Computing and
information Technology
University of Gujrat,
Pakistan

Sana Asif

Faculty of Computing and
information Technology
University of Gujrat ,
Pakistan

Naveed Anwer Butt

Faculty of Computing and
information Technology
University of Gujrat,
Pakistan

Imran Ashraf

Faculty of Computing and
information Technology
University of Gujrat,
Pakistan

ABSTRACT

This research paper combines the data mining with natural language processing to extract the nuggets of knowledge from massive volume of student feedback dataset on faculty performance. The main objective is to compare two renowned association rule mining and sequential pattern mining algorithms namely Apriori and Generalized Sequential Pattern (GSP) mining in the context of extracting frequent features and opinion words. Student feedback data crawled, pre-process and tagged, then convert in tri-model data files. Both algorithms are applied on prepared data through WEKA 3.7.10 (a machine learning tool) to extract the rules. Mined rules are applied on testing files to extract frequent features and opinion words. Evaluated Results show that GSP is more significant to use for textual data mining than Apriori.

Keywords

Opinion Mining, Sentence Level Sentiment Classification, Sequential Pattern Mining, Apriori, Generalized Sequential Pattern, Opinion Words, Frequent Features

1. INTRODUCTION

To make decision about any matter we always ask the other people “what they think”. Faculty is a building block of any educational institute and students are the ambassador of the teachers. Institutes have different parameter to check the performance of their faculty to enhance the educational quality. One is to take feedback from students about teaching faculty. Universities conduct the online teacher performance survey. Students give their feedback (comments) in textual free format to express their reviews. The dataset used in this research is educational student feedback data to extract the frequently commented features along with their opinion words. Three level sentiment classifications (document, sentence and feature level) exist to summarize customer reviews. [1, 2]

Sentence level sentiment classification is one level deeper to document level opinion mining. It extracts such sentences from reviews documents which contains object, noun (just feature words) and adjectives [1]. Polarity about opinionated object is found by counting positive and negative adjective words used to express opinion about product and its features. If positive words are more than negative, opinion about object is positive

otherwise negative, if both are same then opinion is neutral [3]. To extract nuggets of knowledge from huge amount of opinionated text in databases data mining techniques are used [4, 5]. Such techniques directly refer to Artificial Intelligence. Generalization, classification, clustering, association rule mining, data visualization, neural networks, fuzzy logic, Bayesian networks, genetic algorithm, decision tree, multi agent systems, CRISP- DM model, churn prediction, and Case Based Reasoning etc are major data mining techniques used to extract the related knowledge and information [6].

Classification is a popular and frequently used supervised learning data mining technique. It is used to predict the required outcomes on the basis of dependent (predicted attribute) and independent attributes (dataset related properties). A training dataset and a testing datasets are used to create a predicted model. Created model extract rules from training dataset and apply on test corpora to find the model’s accuracy [7]. Prediction hit rate is used as accuracy parameters of extracted rules to make prediction. [8, 9].

Clustering is an unsupervised learning technique used to gather objects of similar characteristics in form of clusters. Each cluster has dissimilar objects to other cluster [10]. In contrast to classification, clustering has no explicit targeted output associated with input. As the known Class label is unknown so, unsupervised learning is about to learn by observation instead of learn by example [8].

Case based reasoning (CBR) is a supervised machine learning technique used to solve the problem on the basis of past similar problems. The past problem’s solutions are stored in CBR repository called Knowledge base or Case base. To solve the new problem, similar old problems solution retrieve from knowledge base. Such method enhance the reusability of solved problems. May be little bit modification make on solution according to the new problem parameters then changed solution stored in case base repository as new instance [4].

Sequential pattern mining framework (SPMF) is used to extract frequent patterns from huge volume of transactional database. About 50 algorithms exists in SPMF like association rule mining (ARM), frequent pattern mining, clustering, association pattern mining, frequent rule mining, classification mining etc. In feedback comments students frequently comment on the teaching methodology, punctuality, behavior towards students, exam marking, lecture preparation etc. In this research we are concerned the extraction of the frequent commented features and

opinion words. So, we select two sequential pattern mining algorithms Apriori (frequent item set) and Generalized Sequential Pattern.

2. RELATED WORK

Several works have been done in the field of data mining to extract frequent, infrequent features and opinion words. To mine the customer reviews on a product [2] proposed unsupervised algorithm. They find frequent features using Apriori algorithm. Chinese WordNet set classifies opinion words in clauses (positive, negative or neutral) to summarize the comments.

To find the strength of opinions Riloff and Wiebe [10] used a new method of syntactic clues. A wide range of adjective words used to find the strength of opinion. The proposed system used to support information analysts in government, commercial, and political domains, who want to automatically track attitudes and sentiments in the news and on-line forums.

Kim and Myaeng [11] used AutoSlog-TS extraction pattern learning algorithm to improve subjectivity classification. To improve the combination of rule-based algorithms and machine learning techniques opinion Analysis performed on Lexical Clues.

Zaki [12] proposed a semi-supervised learning on sentence level sentiment classification. Proposed system efficiently extract the frequent sequences based on highly precise seed rules. They identified opinion holders and polarity of the sentence as Positive, Negative or Neutral.

Riloff and Wiebe [13] presented a system based on conditional probabilities to identify an opinion holder based on an anaphor resolution technique to improve the performance of syntactic features.

Turney [14] proposed a supervised learning method to make difference between subjective and objective sentences by using Naïve Bayes algorithm on only unannotated text as their training data.

Naveed and Ayesha [15] proposed a feature based opinion mining method “Statistical Opinion Analyzer” (SOA). They used Bayesian probability and frequency distribution to extract the polarity of online customer’s reviews. The proposed system helps the new customer and manufacturer in decision making about product by reviewing the results.

Ghorashiet al. [16] used Apriori and H-Mine algorithms to detect frequent features from product reviews dataset. They obtained more efficient results after applying compactness and redundancy pruning.

Lawrence and Pennock [17] used Apriori algorithm to extract frequent features from huge volume of transactional database. They removed irrelevant features after compactness and redundancy pruning and got significant results.

Hemalatha, Krishnan and Hemamathi [18] used Apriori algorithm to extract non-product features and irrelevant opinion words. They considered infrequent features as irrelevant features and removed them from final results. Somprasertsri and Lalitrojwong [19] proposed a dependency and semantic based approach to summarize the polarity of customer reviews about online purchased document. They focused on extracting the relationship between product feature and its nearby opinion word. Experimented results were compared with Adjacent based, Pattern based and Baseline method results. They evaluated that

proposed method achieved 0.15 % recall and 0.11 % f-score better than others. Jian Pei et. al.[20] presented a sequential pattern growth pattern based on projection method to mine the sequential patterns from large transactional database. They integrated PrefixSpan algorithm to achieve efficient computational results.

3. PROBLEM DEFINITION

Specific instances which are aim to extract from given dataset are defined as following

Definition 1: *Product feature*

The physical attributes of an object are called its features. For example class discipline, teaching methodology, punctuality, and student support are the features of a teacher. Features are parsed as noun or noun phrases and are represented as *_NN* or *_NNS*.

Definition 2: *Opinion Word*

The adjective words used to explain the satisfaction level of student about concerning teacher are called opinion words. Parsed sentiment words are represented as *_JJ* or *_JJS*.

Definition 3: *Frequent feature*

Such features which are commonly commented by the reviewers are called frequent features. Frequent features are detected from sentences which have at least one feature word along with its opinion. Frequent features are also called explicit features. Lecture preparation, lecture delivery, involve the students, course related knowledge, punctuality and regularity, exams marking, and material are the frequent feature of a teacher because every student is concerned to it.

4. METHODOLOGY

4.1 Crawl Reviews

Universities Conducted online Teacher evaluation surveys. Their targeted population to conduct survey is students of different universities. Students give their feedback or reviews in comments in free textual format about each teacher separately. In this research five (5) files of online student’s reviews in free format chosen as training and testing Dataset. Three files are used as training and two files are used as testing data. These files are stored in Review Database for pre-processing.

4.2 Data Preprocessing

As data is in free format and extracted from web, lot of irrelevant information like HTML Tags, special characters, fake reviews, spelling mistakes, student’s own information removed from reviews documents to make data easy for further use.

4.3 POS Tagging

As the features along their opinion words have to identified, such sentences are required which contain feature and its corresponding adjectives. To perform this function Part of Speech Tagging is the best choice. In this research online available Go Tagger used to tag the data. Tagger tags the features as noun and opinion words as adjectives.

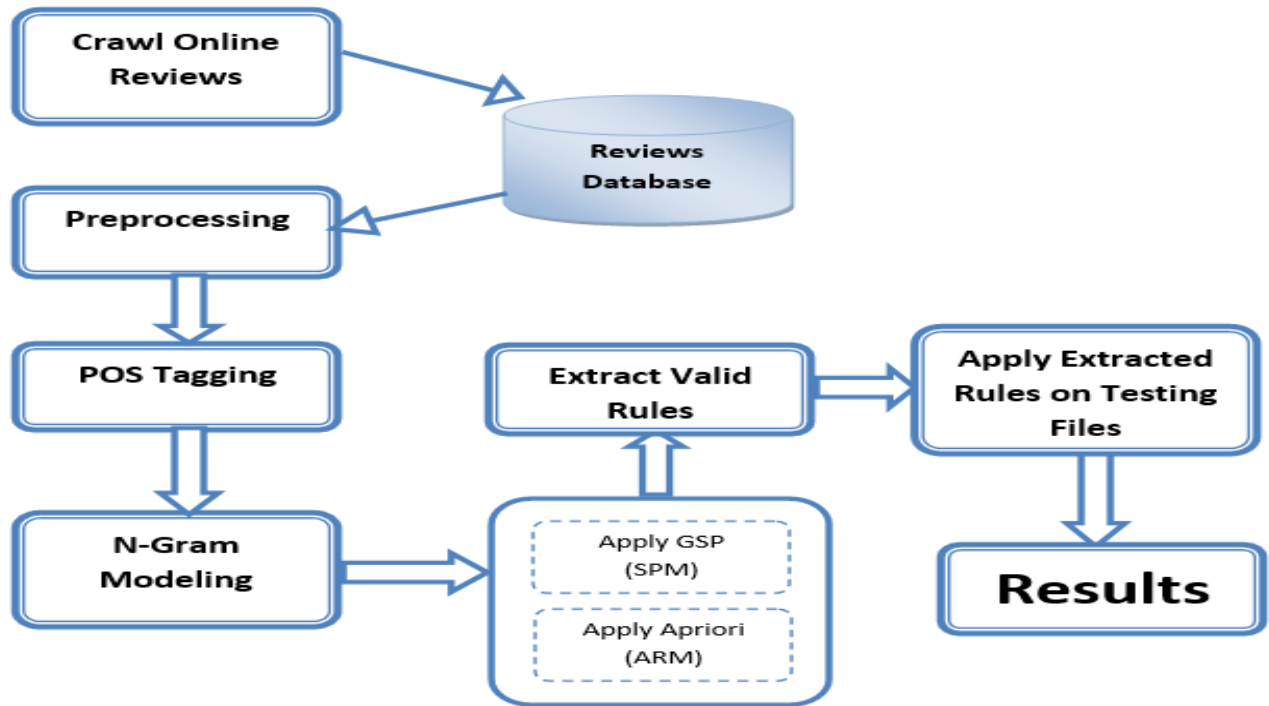


Fig1: Methodology

For example the sentence “his teaching methodology is excellent” will tag as “his_ PRP teaching_ NN methodology_ NN is_ VBZ excellent _JJ”. After POS Tagging all the noun features with tag _NN are replaced with letter “f” to make file more generalized for any Object. This strategy is adopted to find the more general language patterns.

Example:

His_ PRP teaching_ NN methodology_ NN is_ VBZ excellent _JJ

After replacing feature and adjective word, the new sentence will be

his_ PRP F_ NN is_ VBZ A_JJ

Here teaching and methodology both words are nouns so we make them one noun and replaced by F which shows that it is a feature and excellent is replaced by A that show Adjective. Usually all JJ are adjective but all NN (nouns) are not features.

4.4 N Gram Modeling:

“An n-gram is a sub-sequence of n items from a given sequence”. [19]. N-gram modelling is used to convert unstructured data into structure data. The n-gram with size 1 is known as unigram; with size two as bigram, size 3 is as trigram and size 5 is Pentagram.

Sometimes consumers use long sentences or even one line paragraph to comment. Trigram modelling applied on tagged data to split sentences in meaningful form. Using sentence as it is May lose the important information, any feature or adjective word and it may cause to generate fake or bogus rules. Trigram modelling

extracts valid and meaningful rules and stored with .arff extension as WEKA file.

4.5 Apply Sequential Pattern Mining Algorithms

Association rule mining techniques are the best approach to extract valid rules from textual data. The main objective of this research is to extract the feature words and opinion words from customer reviews dataset. Two SPM algorithms Apriori and Generalized Sequential Pattern (GSP) are selected to apply on review documents. Extracted rules are applied on testing data to check whether the rules are applicable or not. Then the best rules are implemented on pre-processed dataset to extract feature words and opinion words. Both algorithms are applied on selected data to find out which one is the best to achieve the goal.

4.5.1 Apriori Algorithm Implementation:

Apriori is an Association Rule Mining algorithm used to extract the valid rules on the basis of association among attributes. Agarwals introduced Apriori algorithm to find the regularities in different item sets from huge volume of transaction database. Support and confidence are the two main parameters used in algorithm to extract the best rules.

$$\text{Support} = P(X \cup Y) / N$$

Where $P(X \cup Y)$ = number of times X and Y appear together

N= total number of items

$$\text{Confidence} = P(X \cup Y) / P(X)$$

Where $P(X \cup Y)$ = number of times X and Y appear together

$P(X)$ = number of times X appears in dataset

Prepared data files are uploaded in machine learning WEKA tool one by one to mine the best rules.

The set parameters are Lower **bound min support = 0.04**
Metric type =lift **Num rule** =100 **Upper bound min support**=0.9.

There are many applications in which some items appear very frequently in the data, while some other items rarely appear. If set the value of support and confidence set high, two problems will encounter

1. If the min support is set too high, the rules that involve infrequent items or rare items in the data will lose.
2. In order to find rules that involve both frequent and rare items, min support have ton set very low.

The said parameters applied on all training files and get rules. To extract features the rules having f_NN and f_NNS are considered to be valid. For example for CANON POWER SHOT SD500 WEKA extracted 75 rules, all the invalid rules are eliminated like $w3=NN\ 104 \implies w2=VB\ 50$. Eliminate repeated as well as such rules whose support value is minimum to 0.04. As rules are to be generalize so eliminate such rules which already existed in another file. For example rule $w2=JJ\ 40 \implies w3=f_NN\ 18$ exists in almost all product file rules. Such rules are eliminated from other files and keep single time. Final best extracted rules are as following

4.5.2 GSP Algorithm Implementation

GSP is sequential pattern mining algorithm based on Apriori algorithm is used for sequence mining. GSP scans the database multiple times; in the first scan all the frequent item sets are extracted which are called candidate 1 (C-1) sequence generation. From C-1 candidate 2 sequence generation set is formed and from frequent item sets of C-3 sequence is generated. This process is repeated until no frequent item remains. Then the valid combinations are counted on the basis of Support parameter.

In this research, prepared trigram model file uploaded in WEKA 3.7.10 and applied GSP algorithm from association rule mining section. Rules are generated up to candidate – 3 sequence generation set. For feature extraction such rules were selected which contained “ f_NN or f_NNS ” and for opinion word extraction picked such rules which contained “ A_JJ or A_JJR ”. LIKE Apriori Validity of rules depends on minimum support value. Repeated rules and less than minimum support values were removed from rules set and best extracted rules are as follows:

Table 2: GSP best Extracted Rules

1-sequences	2-sequences	3-sequences
{DT}	{VBN, f_NN }	{ f_NNP }{NN,PRP}
{NN}	{VBD, f_NN }	{ f_NN }{NN,DT}
{VBZ}	{ f_NNS ,WRB}	{ f_NN }{NN,VBZ}

{IN}	{A_JJ,NN}	{ f_NN ,NN}{DT}
{ f_NNS }	{DT,A_JJ}	{NN_F}{PRP,VBZ}
{ f_NN }	{NN,A_JJ}	{NN_F}{WDT,VBZ}
{A_JJ}	{VBZ,A_JJ}	{NN_F}{WDT}{WDT}
{DT}	{F_NNS,A_JJ}	{A_JJ,PRP}{PRP}
{NN}	{F_NN,VBZ}	{NN_F}{A_JJ,VBZ}
{VBZ}	{VB,A_JJS}	{NN_F}{PRP,VBZ}
{IN}	{NNS,A_JJ}	{A_JJ}{WDT,VBZ}

The one sequence rules excluded as it consist on single parameter and to check the existence of single attribute makes no sense. So 1-candidate sequence generation ignored and best rules are extracted from 2 and 3 candidate generation set.

Table 1: Apriori Best Extracted Rules

F_NNS, IN	NN_F,IN ,DT
DT ,F_NNP,F_NNP	DT ,NN_F,VBZ
IN, F_NNS	DT ,JJ_A,NN_F
F_NNP ,VBZ	DT ,NN_F ,NN_F
F_NNP ,PRP	NN_F ,VBP
F_NNP,F_NNP	NN_F ,RB
VBD ,F_NN	DT ,PRP,VBP
F_NNS ,PRP	A_JJ ,TO ,VB
IN ,F_NNP	F_NN ,TO ,VB
IN ,DT,f_NN	A_JJR ,F_NNS ,IN
DT ,f_NNP	CD ,VBG ,A_JJR
JJ ,f_NN	A_JJR ,TO ,VB
f_NN ,VBZ	RBR ,A_JJ ,F_NN
DT ,f_NN	F_NNS,POS ,PRP
f_NN,CC	F_NNS,TO ,VB
f_NN ,PRP	NN ,A_JJR ,TO
f_NN ,NN	WRB ,DT ,F_NN
A_JJ ,f_NNS	WRB ,PRP\$,F_NN
f_NN ,IN	F_NNS ,JJ ,IN
NN ,f_NNS	A_JJ,F_NN ,VBN
f_NN ,DT	CD ,DT ,F_NN
IN ,f_NN	CD ,A_JJS ,RB
DT ,F_NNP ,NNP	CC ,F_NNS ,F_NNS
F_NN ,NN ,PRP	CC ,A_JJS,F_NNS
F_NNP ,NNP ,IN ,DT	DT ,PRP\$,F_NNS

4.6 Best extracted Rules implementation on Testing Files

The best extracted feature extraction and adjective rules combinations are applied on testing files to extract feature and opinion words respectively. Rules are applied one by one on testing files and for each rule the parameter of confusion matrix is calculated to check accuracy of applied combination. A= True Positive (TP)

B= False Negative (FN)

C= False Positive (FP)

D= True Negative (TN)

Form such files for each review file and adjectives are extracted separately

File 1 Adjective Rule

Table 3: Confusion Matrix Parameter Calculated On Feature Extracted Rules

RULES	Features (A)	Non features (C)
VB,F_NNS	1	1
F_NNS,DT	3	1
F_NNS,CC	4	5
F_NN,CC	5	7
F_NN,VBZ	12	3
F_NN,RB	1	2
IN,F_NNS	1	1
F_NNS,NNS	2	2
F_NN,NN	44	33
DT,F_NN	17	39
PRP\$,F_NN	1	10
F_NN,IN	17	18
PRP\$,F_NNS	2	4
VB,F_NN	2	2
F_NN,VBG	1	5
F_NN,JJ	4	11
F_NN,RB	2	11
F_NN,DT	5	4
VBN,F_NN	0	2
VBD,F_NN	0	2
F_NNS,WRB	1	0
F_NN,DT,NN	6	2
F_NN,VBZ,NN	3	0
VBP,F_NN	3	0
NNS,F_NN	0	2
F_NN,PRP,VBP	0	1
F_NN,WDT,VBZ	1	0
A_JJ,F_NNS	11	8
Total	137	168

Table 4: Adjective Rules Confusion Matrix Parameter

RULES	Features (A)	Non features (C)
A_JJ,F_NNS	11	8
NN,A_JJ	2	18
VBZ,A_JJ	8	1
A_JJ,IN	4	1
A_JJ,PRP	1	1
IN,A_JJ	2	6
A_JJR,NN	0	2
Total	28	37

5. Experimental Results

Extracted rules applied on 5 testing files to extract required attributes. Three parameters precision, recall and accuracy used to measure correctness of mined rules. Evaluated parameters calculated separately for features and opinion words from the metrics as shown in table 2. To calculate precision, recall and accuracy following formulas are used

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ \text{Accuracy} &= \frac{TP+TN}{TP+FP+FN+TN} \end{aligned}$$

Table 5 shows the results to extract frequent features from textual data using GSP and Apriori. As the Apriori's average precision is 0.071 and average recall is 0.704. In its contrast by using GSP average precision is 0.862 and average recall is 0.86. It is obvious from calculated results that GSP is more efficient to use to extract frequent features from customer reviews data.

Table 5: Frequent Feature Extraction Results

Dataset Files	GSP		APRIORI	
	Recall	Precision	Recall	Precision
File 1	0.87	0.87	0.72	0.75
File 2	0.79	0.75	0.63	0.62
File 3	0.91	0.93	0.64	0.64
File 4	0.9	0.9	0.78	0.79
File 5	0.84	0.85	0.75	0.75
Average	0.862	0.86	0.704	0.71

Figure 2 demonstrates the results to extract frequent features from textual unstructured dataset using defined algorithms.

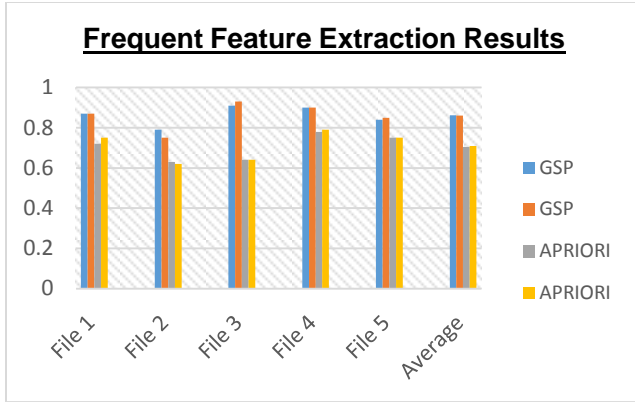


Figure 2: Frequent Feature Extraction Results

Table 6 represents the results to extract opinion words (adjectives) from textual data using GSP and Apriori. As the Apriori's average precision is 0.94 and average recall is 0.952 while by using GSP average precision is 0.956 and average recall is 0.962. It is clear from calculated results that GSP is more efficient to extract Opinion Words from customer reviews data in sentence level text mining.

Table 6: Opinion Word Extraction Results

Dataset Files	GSP		APRIORI	
	Recall	Precision	Recall	Precision
File 1	0.95	0.93	0.96	0.93
File 2	0.98	0.97	0.95	0.92
File 3	0.96	0.95	0.95	0.94
File 4	0.95	0.97	0.94	0.94
File 5	0.97	0.96	0.96	0.97
Average	0.962	0.956	0.952	0.94

Figure 3 demonstrates the results to extract opinion words from customer reviews dataset. Figure shows that GSP is better than Apriori to extract frequent features.

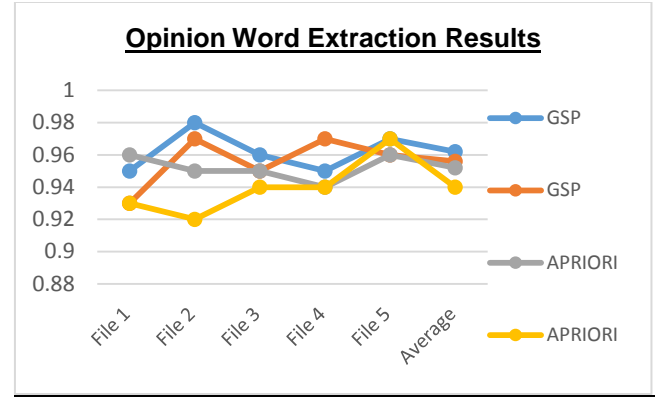


Figure 3: Opinion Word Extraction Results

Table 7 represents the accuracy results to extract frequent features and opinion words using Apriori and GSP. To extract frequent features accuracy using Apriori is 81% while GSP achieved 87 % accuracy. As well as to extract adjective words Apriori's accuracy is 93 % and GSP achieved 95 % accuracy. Results shows that GSP is the better choice to extract nouns and adjectives than Apriori.

Table 7: Accuracy Comparison of Both Algorithms

Dataset Files	GSP Accuracy (%)		Apriori Accuracy (%)	
	Noun	Adjective	Noun	Adjective
File 1	91.45	96.8	82	91.0
File 2	84.77	96.3	80.3	96.56
File 3	91.0	95.56	81.0	92.47
File 4	85.52	93.56	79.0	95.0
File 5	86.75	95.99	85.0	93.0
Average	87.898	95.642	81.46	93.606

Figure 4 represents the accuracy results to extract features and opinion words of both algorithms. Figure shows that GSP is better than Apriori to extract frequent features.

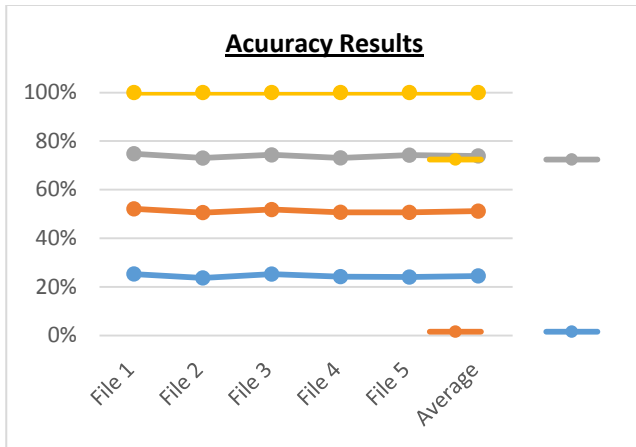


Figure 4: Accuracy Comparison Results

Table 8 represents the precision and recall results to extract infrequent features. Average precision and recall of GSP is 0.6968 and average precision and recall of Apriori is 0.5834. Results clearly shows that GSP is more effective to extract implicit features from textual data.

6. CONCLUSION

In this paper, sentence level opinion mining used to extract the commented frequent features and opinion words from students feedback dataset in textual free format about faculty evaluation. A generalized method is proposed to extract mentioned metrics. Proposed method executes the complete cycle of data mining and applies two Sequential pattern algorithms Generalized sequential pattern mining and Apriori on prepared data to find the most significant algorithm to extract nouns and adjectives. Experiment results show that GSP is 6% more efficient for frequent features and 2.03% for opinion word as compared to Apriori. So GSP is more robust and effective to use than Apriori on textual data to mine the required knowledge.

7. FUTURE WORK

In future, we will use penta-gram model, other association rule mining and machine learning algorithms to find the most optimistic algorithm to extract features and opinionated terms. We also aim at working on strength of opinion words and classify them in positive and negative clauses.

8. REFERENCES

- [1] Harb, A., Plantié, M., Dray, G., Roche, M., Troussset, F. and Poncelet, P., "Web Opinion Mining: How to extract opinions from blogs?", International Conference on Soft Computing as Tran disciplinary Science in 2008.
- [2] Lin, W.H., Wilson,T., Wiebe,J. and Hauptmann, A., "Which Side are You on? Identifying Perspectives at the Document and Sentence Levels", Proceedings of the 10th Conference on Computational Natural Language in 2006.
- [3] B. Seerat and F.Azam "Opinion Mining: Issues and Challenges (A survey)" International Journal of Computer Applications (0975 – 8887) Volume 49– No.9, July 2012.
- [4] A. Aamodt, H. A. Sandtorv, O. M. Winnem "Combining Case Based Reasoning and Data Mining - A way of revealing and reusing RAMS experience" Safety and Reliability; Proceedings of ESREL, Trondheim, June 16-19, 1998.
- [5] A. BALAHUR and A. MONTOYO, "A Feature Dependent Method for Opinion Mining and Classification" Natural Language Processing and Knowledge Engineering, 2008. NLP-KE '08. International Conference on Date of
- [6] Conference: 19-22 Oct. 2008.
- [7] X. WANG, G. HONG FU "Chinese subjectivity detection using a sentiment density-based naive Bayesian classifier" conference on machine learning and cybernetics (icmlc) international in 2010.
- [8] Pang, B. and Lee, L. 2004, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proceedings of the Association for Computational Linguistics ACL 2004.
- [9] X. Su, G. Gao, Y. Tian, "A Framework to Answer Questions of Opinion Type" Seventh Web Information Systems and Applications Conference in 2010.
- [10] Yu,H. and Hatzivassiloglou ,V. 2003,"Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences", Proceedings of EMNLP-03, 8th Conference on Empirical Methods in Natural Language Processing pages 129–136 in 2003.
- [11] Riloff, E., Wiebe, J., and Philips, W., "Exploiting Subjectivity Classification to Improve Information Extraction", Proceedings of the 20th National Conference on Artificial Intelligence in 2005.
- [12] Kim, Y. and Myaeng, S., "Opinion Analysis based on Lexical Clues and their Expansion", Proceedings of NII Test Collection for Information Retrieval in 2007.
- [13] Zaki, M. J. "Efficient enumeration of frequent sequences", in '7th International Conference on Information and Knowledge Management', ACM Press, Bethesda ,Maryland, United States, pp. 68– 75 in 1996.
- [14] Riloff, E. and Wiebe, J. "Learning Extraction Patterns for Subjective Expressions", Proceedings of EMNLP-03, 8th Conference on Empirical Methods in Natural Language in 2003.
- [15] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424, 2002.
- [16] N, Anwer and A, Rashid "Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics" Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on 16-18 Aug. 2010.
- [17] Ghorashi, Ibrahim et al. "A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012

- [18] K. Dave, S. Lawrence and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," in Proceedings of the 12th international conference on World Wide, New York, 2003.
- [19] R. Hemalatha, A. Krishnan and R. Hemamathi, "Mining Frequent Item Sets More Efficiently Using ITL Mining," in 3rd International CALIBER, Ahmedabad, 2005.
- [20] Somprasertsri and Lalitrojwong, "Mining FeatureOpinion in Online Customer Reviews for Opinion Summarization" Journal of Universal Computer Science, vol. 16, no. 6 (2010), 938-955 submitted: 15/9/09, accepted: 4/3/10, appeared: 28/3/10 c J.UCS
- [21] Jian and Behzad et al. "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach" IEEE Transactions On Knowledge And Data Engineering, Vol. 16, No. 10, October 2004
- [22] W. J. Jia, S. Zhang, Y. Ju Xia, J. Zhang and H. Yu "A Novel Product Features Categorize Method based on Twice-Clustering", 2010 International Conference on Web Information Systems and Mining in 2010.
- [23] T. Khushboo "Mining of Sentence Level Opinion Using Supervised Term Weighted Approach of Naïve Bayesian Algorithm", Int. Journal. Computer Technology & Applications, Vol 3 IJCTA | MAYJUNE 2012.
- [24] N. M. Shelke, S. Deshpande and V. Thakre "Survey of Techniques for Opinion Mining" International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.
- [25] D. S. Deshpande "A Survey on Web Data Mining Applications" Emerging Trends in Computer Science and Information Technology -2012 (ETCSIT2012) Proceedings published in International Journal of Computer Applications® (IJCA).