ISSN:0975-9646

Swati V. Gupta et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1362-1368

Multi-Objective Sentiment Analysis Using Evolutionary Algorithm for Mining Positive & Negative Association Rules

Swati V. Gupta[1], Madhuri S. Joshi[2]

[1,2] *Department of Computer Science & Engineering, MGM's Jawaharlal Nehru Engineering College, N-6 CIDCO, Aurangabad - 431003, M. S., India*

Abstract- **Most of the algorithms for mining Quantitative Association Rules (QAR) focuses on positive dependencies without paying particular attention to negative dependencies. The latter may be worth taking into account, however, as they relate the presence of certain items to the absence of others. The algorithms used to extract such rules usually consider only one evaluation criterion in measuring the quality of generated rules. Recently, some researchers have framed the process of extracting association rules as a multiobjective problem, allowing us to jointly optimize several measures that can present different degrees of tradeoff depending on the dataset used. In this paper, a Multi-Objective Sentiment Analysis is done using Evolutionary Algorithm (EA) for mining positive & negative Association rules. It is an important methodological application in the world of Data Mining (DM). This paper includes the approach and technique of Multi-Objective Positive Negative Association Rule (MOPNAR) based *predictive Sentiment Analysis*, which is based on huge dataset of multiple opinion obtained. In this study multiple opinions from customer, data analyst, writers, and composers has been used which are in the form of text for identification of predictive sentiments.**

Index Terms- **MOPNAR, Multi-Objective Sentiment Analysis, Opinion Mining, Sentiment Analysis.**

I. INTRODUCTION

From the last decade, the digital revolution has provided relatively inexpensive and accessible means of collecting and storing data. This unlimited growth of data has led to a situation in which the knowledge extraction process is more difficult and, in most cases, leads to problems of scalability and/or complexity [2]. Association discovery is one of the most common data mining techniques used to extract interesting knowledge from large datasets [3]. Association rules are used to identify and represent dependencies between items in a dataset [4]. Multi-objective sentiment analysis and predictive mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in Natural Language Processing (NLP) and is also widely studied in data mining, Web mining, and text mining. In fact, this research has spread outside of computer science to the management sciences and social sciences due to its importance to business and society as a whole. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. For the first time in human history, now a huge volume of data is available which is known as opinionated data and that data is recorded in digital form for analysis.

Multi-Objective Sentiment analysis systems are being applied in almost every business and social domain because opinions are central to almost all human activities and are key influences of people behaviors. People beliefs and perceptions of reality, and the choices make by the user, are largely conditioned on how others see and evaluate the world. For this reason, when person need to make a decision they often seek out the opinions of others. This is true not only for individuals but also for organizations.

In this paper, focus is on implementation of a new Multi-Objective Evolutionary Algorithm (MOEA) with positive negative Association rule to achieve a new learning of characteristics and a condition selection for each rule, while presenting an exterior population (data) and a resuming process to store all the non- conquered rules found and to recover the variety of the rules. This proposal introduces an Evolutionary Process (EP) and a restarting process to the evolutionary model in order to store all the non-dominated rules found and to improve the variety of the rule set obtained. There are numerous methods to changing the multi-objective optimization problem into a number of scalar optimization problems. This produces an optimized approach to understand from a user's standpoint, and with extraordinary values for the interestingness events in all information contents.

II. PRELIMINARY: POSITIVE AND NEGATIVE QUANTITATIVE ASSOCIATION RULES

Many previous studies for mining association rules have focused on datasets with binary or discrete values; however, the data in real-world applications usually consists of quantitative values. The association rules obtained from datasets with quantitative values is known as QARs, where each item is a pair attribute-interval [7]. For instance, a positive QAR could be Age $\in$ {30, 52} and Salary $\in$ {3000, 3500} $\rightarrow$ NumCars $\in$ {3, 4}.

Most of these algorithms have only focused on positive rules, i.e., only those item sets appearing frequently together will be discovered. However, the negative association rules may also be interesting as they offer information that could be used to support decisions for applications. Negative association rules consider the same sets of items as positive association rules but may also include negated items within the antecedent ($\neg X \to Y$) or the consequent ($X \to \neg Y$) or both ($\neg X \to \neg Y$) [8]. For instance, a simple example of a negative QAR is Weight ∈ ¬ {15, 30} and Height ∈ {90, 150} → Age ∈ ¬ {4, 28}.

Notice that, positive association rules only include positive items whereas negative association rules include at least one negative item. Fig. 1 shows the domain of the positive item *Height* ∈ {90, 150} and the negative item *Age* ∈ ¬ {4, 28}. Support and confidence are the most common measures used to assess QARs, both of them based on the support of an item set. The support of the item set I is defined as

$$SUP(I) = |\{e \in D \mid I \in e\}| / |D| \qquad (1)$$

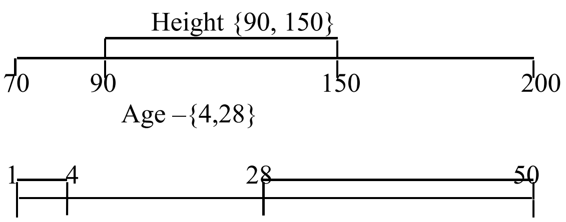


Figure 1: Examples of positive and negative item

where the numerator is the number of examples in the dataset *D* covered by the item set *I*, and | *D* | is the number of examples in the dataset. Thus, the support and confidence for a rule $X \to Y$ are defined as

$$support(X \to Y) = SUP(XY) \qquad (2)$$

$$confidence(X \to Y) = SUP(XY)/SUP(X) \qquad (3)$$

The classic techniques for mining association rules attempt to discover rules whose support and confidence are greater than the user-defined threshold's minimum support (minSup) and minimum confidence (minConf). However, several authors have noted some drawbacks of this framework that lead it to find many misleading rules [8], [9], [10]. On one hand, the confidence measure does not detect statistical independence or negative dependence between items; because it does not take into account the consequent support. On the other hand, item sets with very high support are a source of misleading rules because they exist in most of the examples and therefore any item set may seem to be a good predictor of the presence of the high-support item set.

In recent years, several researchers have proposed other measures for the selection and ranking of examples according to their potential interest to the user [11], [12]. Here the brief description of some of those that have been used in this paper.

The conviction measure analyzes the dependence between X and $\neg Y$, where $\neg Y$ means the absence of Y [10]. Its domain is {0,∞}, where values less than one represent negative dependence, a value of one represents independence, and values higher than one represent positive dependence. The main drawbacks of this measure are that it is difficult to define a conviction threshold because its range is not bounded, and this measure does not decrease when the support of the antecedent increases and the rest of the parameters remain the same. Conviction for a rule $X \to Y$ is defined as

$$conviction(X \to Y) = (SUP(X)SUP(\neg Y))/SUP(X\neg Y) \qquad (4)$$

Notice that this measure obtains an undefined value (NAN) when SUP(Y)=1. In this case, programmer will consider the conviction value to be one, because it denotes independence.

The lift measure represents the ratio between the confidence of the rule and the expected confidence of the rule [13]. As with conviction, its domain is {0, ∞}, where values less than one imply negative dependence, one implies independence, and values higher than one imply positive dependence. The main drawback of this measure is that it is difficult to define a lift threshold because its range is not bounded. Lift for a rule $X \to Y$ is defined as

$$lift(X \to Y) = SUP(XY)/(SUP(X)SUP(Y)) \qquad (5)$$

The Certainty Factor (CF) is interpreted as a measure of variation of the probability that Y is in a transaction when programmer consider only those transactions where X is present [14]. Its domain is {-1,1}, where values less than zero represent negative dependence, zero represents independence, and values higher than zero represent positive dependence. This measure for a rule $X \to Y$ is defined in three ways depending on whether the confidence is less than, greater or equal to $SUP(Y)$

if $confidence(X \to Y) > SUP(Y)$

$$(confidence(X \to Y) - SUP(Y))/(1 - SUP(Y)) \qquad (6)$$

if $confidence(X \to Y) < SUP(Y)$

$$(confidence(X \to Y) - SUP(Y))/SUP(Y) \qquad (7)$$

Otherwise is 0.

The netconf measure evaluates the rule based on the support of the rule and its antecedent and consequent support [15]. Netconf obtains values in {-1, 1}, where positive values represent positive dependence, negative values represent negative dependence, and zero represents independence. Netconf for a rule $X \to Y$ is defined as

$$netconf\,(X \to Y) = SUP(XY) - SUP(X)SUP(Y)/(SUP(X)(1 - SUP(X))) \qquad (8)$$

Notice that if this measure obtains NAN programmer will consider the nefconf value to be zero, because it denotes independence.

Finally, the yule'sQ measure represents the correlation between two possibly related dichotomous events [16]. This measure takes on values in {-1, 1} where one implies a perfect positive correlation, −1 implies a perfect negative correlation, and zero implies that there is no correlation. This measure satisfies almost all the properties for interesting measures that have been proposed in the literature [10], [11]. Notice that as netconf, if this measure obtains NAN consider there is to be no correlation. Yule's Q for a rule $X \rightarrow Y$ is defined as

$$(SUP(XY)SUP(\neg X\neg Y) - SUP(X\neg Y)SUP(\neg XY))/$$
$$(SUP(XY)SUP(\neg X\neg Y) + SUP(X\neg Y)SUP(\neg XY)) \quad (9)$$

III. LITERATURE SURVEY

Natural Language Processing is a domain of computer science and scientific study of human language i.e. linguistics which is related with the interaction or interface between the human (natural) language and computer [17]. Opinion mining or Sentiment analysis refers to a broad area of Natural Language Processing and text mining. It is concern not with the topic a document is about but with opinion it expresses hat is the aim is to determine the attitude (feeling, emotion and subjectivities) of a speaker or writer with respect to some topic to determine opinion polarity. Initially it was applied for classifying a movie as good or bad based on positive or negative opinion. Later it expanded to star rating predictions, product reviews travel advice and other decision making processes.

According to the survey performed by Bo Pang and Lillian Lee, Sentiment analysis identifies the viewpoints of a text. For example, classifying a movie review as thumbs up (recommended) or thumbs down (not recommended). Previous methods focused on selective lexical features (e.g. word "Good"), then classifying document according to the number of such features that occur anywhere within it. But in contrast later following process were followed:

• Identify the sentences in the given input text as subjective or objective.
• Select and apply a standard machine learning classifier to the extracted result.

The previous technique of Sentiment analysis using polarity classification of textual data, estimate the percentage of positivity or negativity of input text by first tagging all the adjectives, adverbs using a POS (Part of Speech) tagger (Marks words in the input text corresponding to a particular part of speech). Also estimate the positivity or the negativity of the extracted adjectives corresponding to its value in the SentiWordNet (derived from WordNet, a lexical database, where numerical value indicating polarity sentiment, i.e. positive or negative, information corresponds to each word in it). In order to estimate sentiment orientation they count the positive and negative terms values. Finally, they assign estimated polarity to the given corpus. The history of the phrase sentiment analysis parallels that of "opinion mining" in certain respects. The term "sentiment" used in reference to the automatic analysis of evaluative text and tracking of the predictive judgments that appears in 2001 paper by Das and Chen [19]. Subsequently, this concept was adopted and enhanced by Turney [20]. In the following year, the concept was carried on by Nasukawa & Yi. These events together may explain the popularity of "sentiment analysis" among communities self-identified as focused on NLP.

A sizeable number of papers mentioning "sentiment analysis" focus on the specific application of classifying customer review as to their polarity "positive" or "negative". Sentiment analysis are extensively studied at different levels such as document level, sentence level, and attribute or feature level. Further details about these levels are presented in the following sub-sections.

IV. PROPOSED SYSTEM

This section provides proposed predictive sentiment classification system using Multi-Objective Positive & Negative Association Rules. The use of this technique for specific segregation of the dataset consist a sequence of steps which are explained as below. The flow chart of the proposed methodology is given in figure 2.

The overall sequence of the proposed methodology is given as follows:

1. Insertion of data with multi-objective essence.
2. Identification of valid useful sentence & its segmentation.
3. Tokenization of segmented sentences.
4. Segregation of the tokenized part to obtain multi-objective positive & negative sentiments hidden in the statements.
5. Efficiency Assessment with Association Rules to gain efficient results of overall analysis.

1. Multi-Objective Data Insertion

For Multi-Objective sentiment analysis programmer choose such kind of data which consists of all kind of hidden sentiments like positive, negative, neutral, exclamatory, questionnaire, etc. This kind of data can be inserted in three ways via manual, automatic file selection or via database storage. Pollution dataset is selected from web reviews.

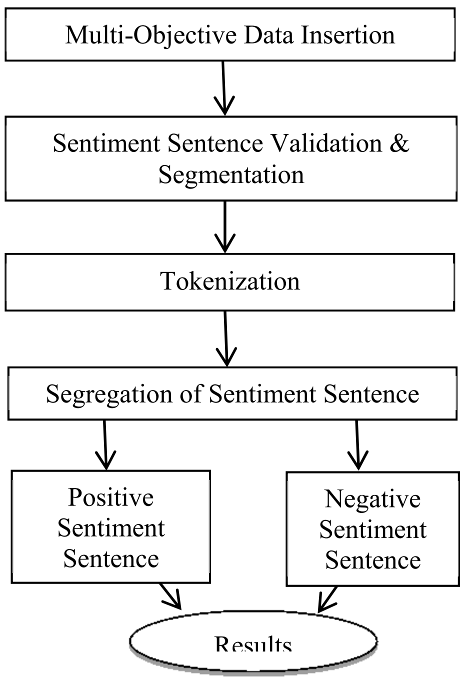


Figure 2: Flow Chart of Proposed Methodology

2. Sentiment Sentence Validation & Segmentation

In this phase, validation of the data is done whether it is in specific desired format or not. Also if the huge data in paragraph form then segmentation of paragraph into sentence is performed. Once the entire paragraph is segmented then it will be ready for further analysis.

3. Tokenization

This is the phase where the specific words, punctuations, exclamation, etc. are captured & create token for each of them. In this step programmer also remove the stop words like punctuations which help us for further segregation process.

4. Segregation of Sentiment Sentence

Here the comparison of available sentences with evolutionary algorithm & grammatical rules for sentiment analysis is performed. In this way, segregation of the tokenized part to obtain multi-objective positive & negative sentiments hidden in the statements is accomplished.

5. Efficiency assessment with Association Rule and its result

Once they ready with positive and negative sentiments then apply the MOEA rule focusing on negative sentiments. With the help of association rules following results will comes out which conclude more efficiency compared to existing algorithm.

Input:

N: number of sentences in the form of paragraph; x={x1, x2, x3… xN}.

Output:

1) Positive, negative and neutral sentences are separated.
2) Positive, negative and neutral sentence's count.
3) Values for different parameters.

Steps:

1) Upload input data file containing 'n' number of paragraphs.
2) Data will be validated using sentiment sentence validation.
3) Sentences will be segmented.
4) Tokens will be separated from all sentences.
5) Process of segregation is performed for finding the positive, negative and neutral words by matching the words with POS tagging words.
6) Calculate all parameters values using their formulas.

V. Experimental Analysis

Experiments have been carried out to analyze the performance of this newly proposed algorithm. There related topics are as follows:

1. Dataset

In order to analyze the performance of the proposed approach, consider two different types of dataset, one of which is related to the pollution and another is movie reviews. Datasets used in this project is in the form of text, known as textual data. Such textual data contains number of sentences, which are the reviews given by the people on the pollution in first dataset and reviews on the movie in second dataset. For analyzing the values of different parameters, variable size of dataset files are used, means different file contains different number of sentences or reviews. So that totally ten files are used, having different number of sentences related to first dataset, named as F7, F14, F21, …, F70 and for second dataset files, naming is F7, F15 and F21.

2. Algorithm considered for comparison is MOPNAR

For obtaining a reduced set of PNQARs with a good trade-off between the number of rules, support and coverage, considering three objectives: comprehensibility, interestingness, and performance. The MOPNAR has extended the multi-objective evolutionary algorithm with MOEA/D-DE algorithm in order to perform an evolutionary learning of the rules and introduces two new components to its evolutionary model: an EP and a restarting process [22].

The MOEA was extended based on decomposition (MOEA/D, MOEA/D-DE, which decomposes the multiobjective optimization problem into N scalar optimization subproblems and uses an EA to optimize these subproblems simultaneously [22]. In order to store all the nondominated rules found, provoke diversity in the population, and improve the coverage of the datasets, they was introduced an EP and a restarting process to the evolutionary model of this MOEA. The EP will keep all the nondominated rules found and updated with the newly generated offspring for each solution of the population. The redundant nondominated rules removed from EP in order to avoid the overlapping rules. A rule is considered redundant if the intervals of all its variables are contained within the intervals of the variables of another rule.

The size of the EP is not limited, which allows us to:
1) Obtain a larger number of rules of the Pareto front regardless of the size of the population;
2) Reduce the size of the population, following a dataset independent approach.

However, the EP will usually contain a reduced set of rules because the non-dominance criteria allow us to maintain only the rules of the Pareto front and that the redundant rules are removed.

To move away from local optima and provoke diversity in the population, the restarting process will be applied when the number of new individuals of the population in one generation is less than α % of the size of the current population n (with α determined by the user, usually at 5%).

3. Parameters considered for comparison

- Average Support, Average Confidence
- Average Lift, Average Certainty Factor
- Net Confidence, Yules Q
- Percentage of Transactions

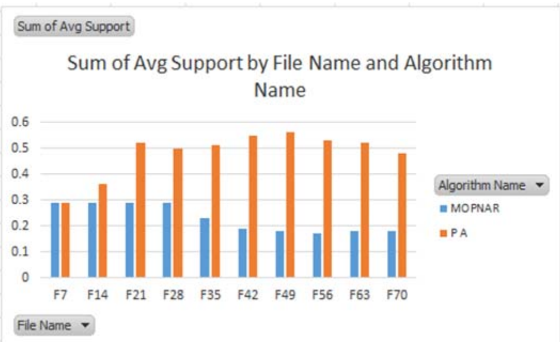
Table1. Result for all Pollution Review Datasets in the comparison with MOPNAR

File Name	Algorithm Name	Avg Support	Avg Confidence	Avg Lift	Avg CF	Net Confidence	YulesQ	%Transactions
F7	MOPNAR	0.29	0.67	2.33	0.53	0.67	1	99
F7	P A	0.29	1	2.33	1	0.8	1	100
F14	MOPNAR	0.29	0.8	2.8	0.72	0.8	1	99
F14	P A	0.36	1.25	3.5	1.39	1.25	1	100
F21	MOPNAR	0.29	1.5	5.25	1.7	1.5	1	99
F21	P A	0.52	1.83	9.63	2.03	2.3	0.75	100
F28	MOPNAR	0.29	1.33	4.67	1.47	1.33	1	99
F28	P A	0.5	1.75	8.17	1.95	2.15	0.78	100
F35	MOPNAR	0.23	0.89	3.89	0.86	0.89	1	99
F35	P A	0.51	2.25	8.75	2.68	2.58	0.76	100
F42	MOPNAR	0.19	0.73	3.82	0.66	0.73	1	99
F42	P A	0.55	2.88	10.98	3.54	3.23	0.71	100
F49	MOPNAR	0.18	0.69	3.85	0.62	0.69	1	99
F49	P A	0.56	3.11	11.97	3.85	3.48	0.69	100
F56	MOPNAR	0.17	0.59	3.41	0.5	0.59	1	99
F56	P A	0.53	3.1	10.58	3.97	3.39	0.73	100
F63	MOPNAR	0.18	0.6	3.3	0.51	0.6	1	99
F63	P A	0.52	2.83	9.35	3.63	3.09	0.76	100
F70	MOPNAR	0.18	0.52	2.92	0.42	0.52	1	99
F70	P A	0.48	2.69	7.86	3.57	2.86	0.84	100

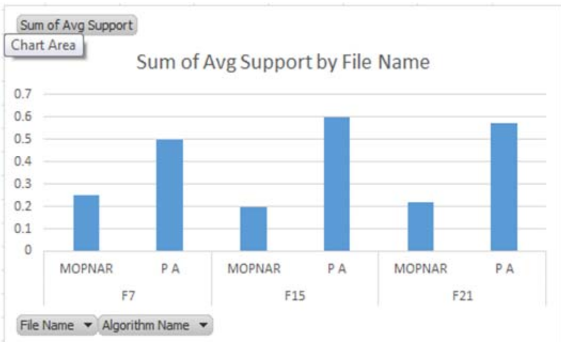
Table 2. Result for all Movie Review Dataset in the comparison with MOPNAR

File Name	Algorithm Name	Avg Support	Avg Confidence	Avg Lift	Avg CF	Net Confidence	YulesQ	%Transactions
F7	MOPNAR	0.25	1	4	1	1	1	99
F7	P A	0.5	2	8	2.33	2.33	0.78	100
F15	MOPNAR	0.2	1	5	1	1	1	99
F15	P A	0.6	3	15	3.5	3.5	0.64	100
F21	MOPNAR	0.22	1	4.6	1	1	1	99
F21	P A	0.57	2.6	11.96	3.04	3.04	0.68	100

Comparison of proposed algorithm with existing MOPNAR using two different datasets in the form of graph

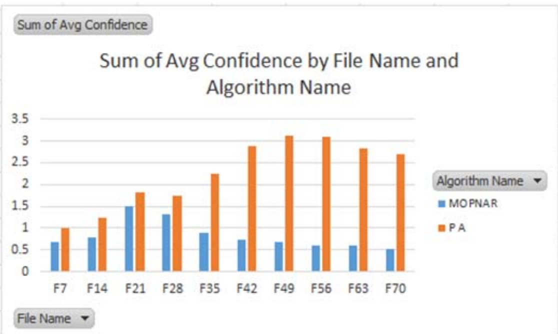


(a) For Pollution Reviews Dataset

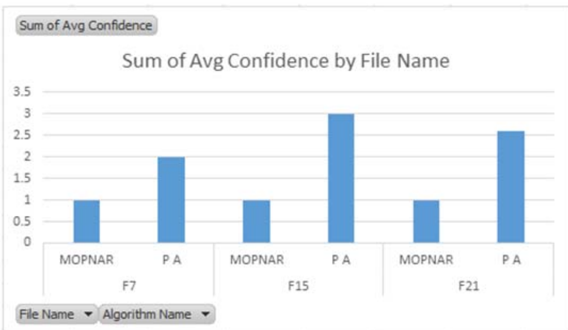


(b) For Movie Review Dataset

Figure 3. Comparison of Average Support parameter of both algorithms

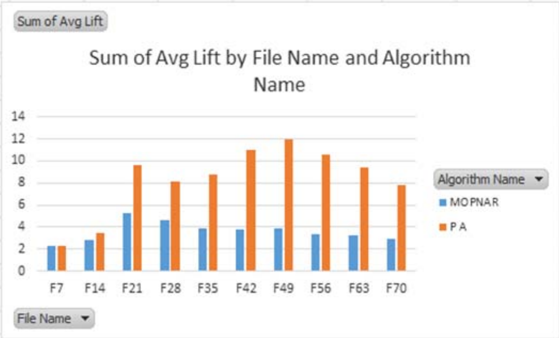


(a) For Pollution Review Dataset

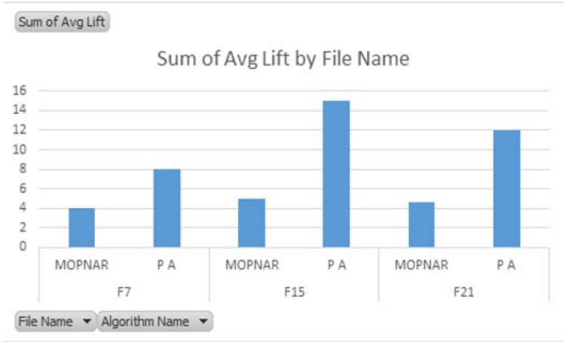


(b) For Movie Review Dataset

Figure 4. Comparison of Average Confidence parameter of both algorithms

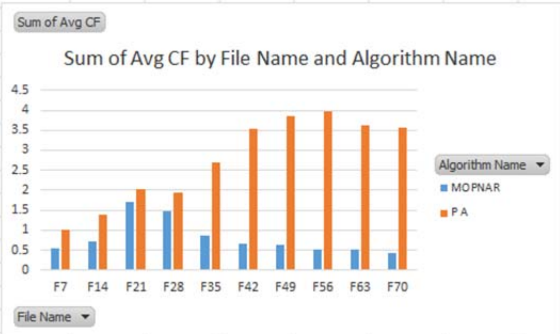


(a) For pollution Review Dataset

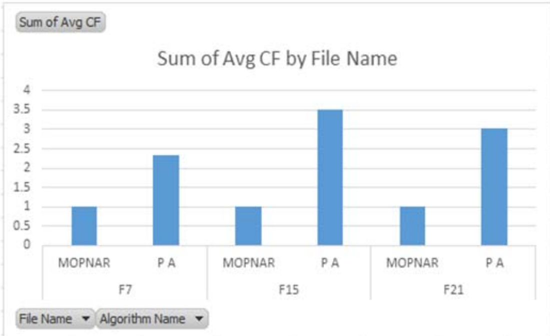


(b) For Movie Review Dataset

Figure 5. Comparison of Average Lift parameter of both algorithms

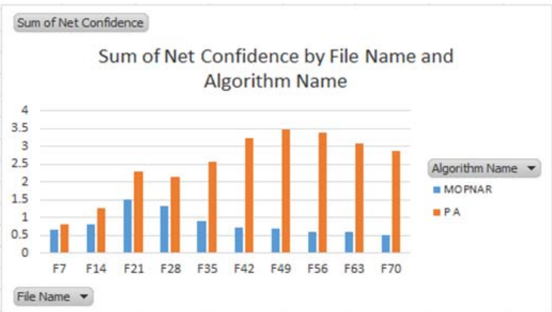


(a) For Pollution Review Dataset

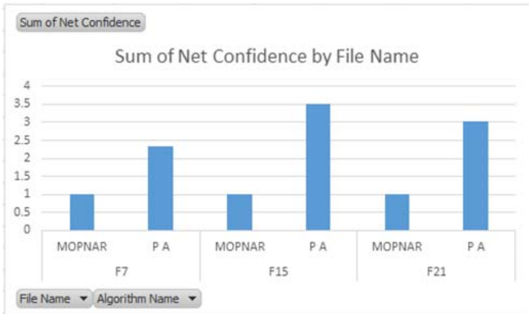


(b) For Movie Review Dataset

Figure 6. Comparison of Certainty Factor parameter of both algorithms

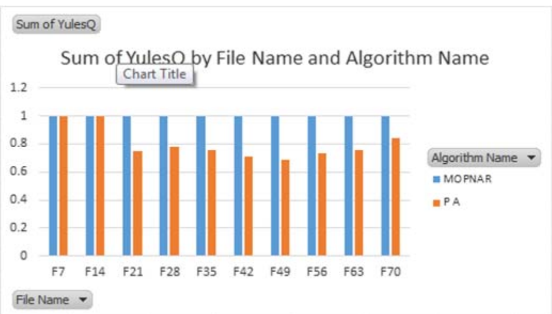


(a) For Pollution Review Dataset

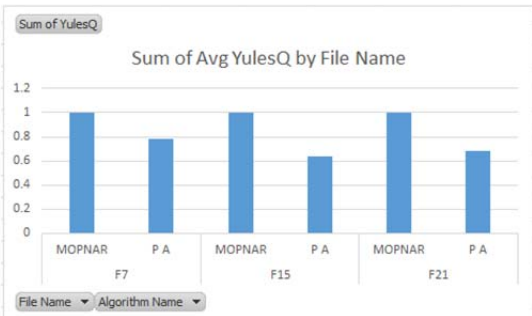


(b) For Movie Review Dataset

Figure 7. Comparison of Net Confidence parameter of both algorithms



(a) For Pollution Review Dataset



(b) For Movie Review Dataset

Figure 8. Comparison of Yules Q parameter of both algorithms

4. Analysis

In this paper, one existing algorithm is used for comparing and analyzing the performance of newly proposed algorithm. The textual data file is used as an input data file, which is uploaded at runtime. The existing algorithm MOPNAR, only concentrated on the positive quantitative association rules, but this proposed algorithm will work on positive as well as negative association rules. Previously developed algorithms only work on the discrete and binary data, but this new algorithm is work on the textual data which will contains the sentiment sentences. These sentiment sentences are the reviews from the public, which contains the sentiment or opinion of peoples. In this paper there are two types of datasets reviews are used one of them is on pollution and second is on movie. Here values of different standard formulas are calculated and compared. Our proposed algorithm gives greater value shown in above table1 and table 2 for all parameters and comparison using graph is also shown in above figures. The proposed algorithm is gone through all transactions present in the input data files of dataset. These graph shows our algorithm is better as compared to existing algorithm.

VI. Conclusion

In this paper, the whole focus is on implementation of a new multi-objective evolutionary algorithm with positive negative Association rule. This proposed algorithm is working on textual data, which is stored in file. The proposed algorithm is compared with MOPNAR algorithm based on the parameters. After comparing this algorithm, on two different datasets, it is found that the proposed algorithm performs better. Proposed algorithm will process all transactions present in input data file and the values calculated for all parameters are giving better values. This algorithm is working on textual data having three different kinds of sentences such as positive, negative and neutral, but previous algorithm is working on the binary and discrete values.

References

[1] Diana Martin, Alejandro Rosete, Jesus Alcala-Fdez, Francisco Herrera, “A New Multiobjective Evolutionary Algorithm for Mining a Reduced Set of Interesting Positive and Negative Quantitative Association Rules”, IEEE Transactions on Evolutionary Computation, vol. 18, no. 1, February 2014.

[2] R. Renesse, K. Birman, and W. Vogels, “Astrolabe: A robust and scalable technology for distributed system monitoring, management, and data mining,” Association for Computing Machinery (ACM) Transactions on Computer Systems, vol. 21, no. 2, pp. 164–206, 2003.

[3] J. Han and M. Kamber, “Data Mining: Concepts and Techniques”, 2nd ed. Burlington, MA, USA: Morgan Kaufmann, 2006.

[4] C. Zhang and S. Zhang, “Association rule mining: Models and algorithms,” in Lecture Notes Computer Science (Lecture Notes in Artificial Intelligence 2307). Berlin/Heidelberg, Germany: Springer-Verlag, 2002.

[5] R. Agrawal, T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases,” in Proc. Special Interest Group on Management of Data (SIGMOD), 1993, pp. 207–216.

[6] R. Agrawal and R. Srikant, “Fast algorithms formining association rules,” in Proceedings International Conferences on Large Data Bases, 1994, pp. 487–499.

[7] R. Srikant and R. Agrawal, “Mining quantitative association rules in large relational tables,” in Proc. Association for Computing Machinery (ACM) Special Interest Group on Management of Data (SIGMOD), 1996, pp. 1–12.

[8] C. Silverstein, S. Brin, and R. Motwani, “Beyond market baskets: Generalizing association rules to dependence rules,” Data Mining Knowledge Discovery, vol. 2, no. 1, pp. 39–68, 1998.

[9] F. Berzal, I. Blanco, D. Sanchez, and M. Vila, “Measuring the accuracy and interest of association rules: A new framework,” Intelligent, Data Analysis, vol. 6, no. 3, pp. 221–235, 2002.

[10] S. Brin, R. Motwani, J. Ullman, and S. Tsur, “Dynamic itemset counting and implication rules for market basket data,” Association for Computing Machinery (ACM) Special Interest Group on Management of Data (SIGMOD) Rec., vol. 26, no. 2, pp. 255–264, 1997.

[11] L. Geng and H. Hamilton, “Interestingness measures for data mining: A survey,” Association for Computing Machinery (ACM) Computer, Surveys, vol. 38, no. 3, pp. 1–32, 2006. Martin et al.: New Multiobjective Evolutionary Algorithm 69.

[12] G. Piatetsky-Shapiro, “Discovery, analysis and resentation of strong rules,” in Knowledge Discovery in Databases. Cambridge, MA, USA: MIT Press, 1991, pp. 229–248.

[13] S. Ramaswamy, S. Mahajan, and A. Silberschatz, “On the discovery of interesting patterns in association rules,” in Proceedings 24th International Conference on Very Large Data Bases, 1998, pp. 368–379.

[14] E. Shortliffe and B. Buchanan, “A model of inexact reasoning in medicine,” Mathematical Biosciences, vol. 23, numbers 3–4, pp. 351–379, 1975.

[15] K. I. Ahn and J. Y. Kim, “Efficient mining of frequent itemsets and a measure of interest for association rule mining,” Journal of Information Knowledge Management, vol. 3, no. 3, pp. 245–257, 2004.

[16] P. Tan, V. Kumar, and J. Srivastava, “Selecting the right interestingness measure for association patterns,” in Proc. 8th Int. Conf. Knowledge discovery in data mining (KDD), 2002, pp. 32–41.

[17] Aparna Trivedi, Apurva Srivastava, Ingita Singh, Karishma Singh and Suneet Kumar Gupta “Literature Survey on Design and Implementation of Processing Model for Polarity Identification on Textual Data of English Language”.

[18] Md. Daiyan1, Dr. S. K. Tiwari, “A Literature Review on Opinion Mining and Sentiment Analysis.”

[19] S. Das and M. Chen, “Yahoo! for Amazon: Extracting market sentiment from stock message boards”, in Proceedings of the Asia Pacific Finance Association annual Conference (APFA), 2001.

[20] P. Turney, “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews”, Association for Computational Linguistics (ACL)’02, Philadelphia, Pennsylvania, 2002, pp. 417 - 424.

[21] W.W. Cohen “Text categorization and relational learning”, Proceedings of Machine Learning '95 Conference, Lake Tahoe, CA, Morgan Kaufmann, 1995, pp. 124-132.

[22] H. Li and Q. Zhang, “Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II,” IEEE Transactions on Evolutionary Computation., vol. 13, no. 2, pp. 284–302, April 2009.