Feature Selection Algorithm for enhancing Modeling Efficiency

Manal Mostafa
Computer & System Engineering
Al- Azhar University
Cairo – Egypt

Mohamed Gamal
Computer & System Engineering
Al- Azhar University
Cairo – Egypt

ABSTRACT

This paper presents High Probability Minimum Redundancy, HPMR, as a new algorithm for employing the most predictive features to contribute dimensionality reduction. The proposed algorithm is useful for finding new, optimal, and more informative features maintaining acceptable classification accuracy. A problem encountered in many large-scale information applications relevant to expert and intelligent systems such as pattern recognition, bioinformatics, social media content classification where data sets containing massive numbers of features. Implementing categorization on these huge, uneven, useless datasets with the overwhelming number of features is just a waste of time degrading the efficiency of classification algorithms and hence the results are not much accurate. HPMR controls the tradeoff between relevance and redundancy by selecting new feature subset that retains sufficient information to discriminate well among classes.

To emphasize the significance of HPMR, it has been relied upon to develop an intelligent system for Arabic sentiment analysis on social media. Additionally, the performance of such algorithm is quantitatively compared with other traditional dimensionality reduction techniques in terms of performance accuracy, dataset reduction percentage, training time. Experimental results prove that HPMR cannot only diminish the feature vector but also can significantly enhance the performance of the well-known classifiers.

General Terms

Dimension Reduction, Text Analysis, Machine Learning, Classification.

Keywords

Feature Selection, HPMR, Chi-squared, Social Media, Arabic, SA, SVM.

1. INTRODUCTION

Feature Selection (FS) has become increasingly significant in a wide range of scientific disciplines [33]. In this paper, FS is considered for the purpose of Sentiment Analysis (SA). For social media SA, the feature vector dimensionality is usually huge [4] [17]. Such high dimensionality can be a severe obstacle for classification algorithms [20]. Most Machine Learning (ML) and data mining techniques may not be effective for high-dimensional data [29]. Curse of dimensionality [10] [21] [26], query accuracy and efficiency degrade rapidly as the dimensions increase. Moreover, the high dimensionality of text makes the analysis computationally expensive and affects the generalization capability of the classifier [13] [17] [28]. These difficulties can be alleviated by applying FS approaches before performing SA tasks [4] [18]. Actually, using FS helps to improve accuracy, center on a key feature space of sentiment

discriminators and tends to diminish over-fitting [12]. It also provides insight into the more informative features for each sentiment class [1]. Besides these keys factors, a reduction of the feature space leads to predictive insights into the nature of the problem under investigation [5] [33], helps to interpret the model by determining which factors really influence the output to be predicted, and simplifies using different visualization techniques [26]. The mathematical procedures making a possible reduction are called Dimensionality Reduction (DR) techniques. Those techniques generally can be categorized into Feature Extraction (FE), and FS approaches. Actually, FE approaches are more efficient than the FS techniques [20] [26]. FE algorithms attempt to select more informative features by projecting the initial feature space into a lower-dimensional one through some algebraic transformations such as Principal Component Analysis (PCA) [7] and Linear Discriminate Analysis (LDA) [30]. Unlike FE, FS algorithms aim at finding out a subset of the most representative features according to some criteria [15] i.e., the features are ranked according to their individual predictive power. Various FS methods have been proposed for selecting dissimilar feature sets such as Information Gain (IG) [20] X^2-test (CHI) [31], and Mutual Information (MI).

All aforementioned algorithms reduce the dimensionality of feature space and remove redundant, irrelevant, or noisy data from the corpus [23]. Efficiently, it brings the immediate effects for application, speeding up a classification algorithm [15], improving the quality of the feature space and thereof the performance of data mining, and increasing the comprehensibility of the mining results. However, the complexity of these approaches is still high [22].

Substantially, HPMR tries to find an optimal subset of attributes which meets the following criteria: (i) appropriately, reduces the cost of feature measurement, (ii) enhances the classification preciseness and makes the patterns easier to understand (iii) allows higher classification accuracy and, (iv) must not lead to significant loss of information. This technique is employed in combination with SVM and the results are compared with some conventional FS techniques, specifically, Chi statistics. By reducing the number of features considered by a classifier, both techniques can improve classification speed and reduce the memory issues.

To examine the applicability of HPMR, it has been used to develop an intelligent prototype to perform SA for Arabic social media. Consequently, the performance of such system is quantitatively compared with some conventional DR techniques. The comparison as such is based on dataset reduction percentage, training time, and performance accuracy.

The rest of this paper is organized as follows: Related work was surveyed in section 2 with a focus on FS methods for SA. Section 3 introduces the proposed HPMR, elaborates the

corresponding framework of SA for social media and discusses the merits of using HPMR. Chi test was covered in section 4. Experimental results and a quantitative assessment of HPMR and Chi algorithm are explained in section 5. Section 6 concludes the paper and gives possible directions for future work.

2. RELATED WORK

An exhaustive overview of the state-of-the-art techniques in this field has been established by many authors like [16] [23] [24]. For classification domain, FS techniques can be categorized into three broad classes; Filter, Wrapper and Embedded methods [16] [23]. Ladha and Deepa [24] evaluated several fundamental FS algorithms found in the literature, assessed their performance in a controlled scenario and presented an empirical comparison of these algorithms. Various FS methods, terms, limitations, advantages and available recent innovation in this field had been projected by [23]. In other work [2], Ahmed et al proposed a method for unsupervised FS. They first defined an effective criterion for unsupervised FS which measures the reconstruction error of the data matrix and then presented a novel algorithm based on a recursive formula for minimizing the reconstruction error based on the new subset.

Varieties of FS approaches have been proposed by multiple studies in order to diminish the feature vector in sentiment classification for improved performance of ML methods. The authors of [18] and [4] investigate the applicability of FS methods for SA. The study of [18] utilized ten filter-based FS techniques and compared them against using no FS. Their experiment shows that FS methods can be very effective to alleviate problems associated with high-dimensional data. Choice of suitable ranker and subset size has a significant impact on the classifier performance. In [4], different FS algorithms and sentiment feature lexicons are investigated on movie reviews corpus with a size of 2000 documents. Their experiments show that performance of the classifier depends on an appropriate number of the representative feature selected from the text. Jiguang et al [17] proposed a new FS method based on matrix factorization framework for sentiment FS. The authors observed that the words that are frequent in one category and seldom appear in the other are a candidate to be strong inter-sentiment categorization. The study made by [32] introduces an enhanced lexicon to perform SA for an online product review. The method combines ML and semantic orientation techniques into one framework. Abbasi et al [1] develop sentiment classification approach to English and Arabic Web discourse at the document level. For this purpose, they provide different feature sets consisting of syntactic and stylistic features and develop the entropy weighted genetic algorithm (EWGA) for FS. Their experiment shows that using syntactic and stylistic features with SVM achieves a higher accuracy. However, they avoided semantic features because they are language dependent and need lexicon resource.

However, the classical batch algorithms no longer satisfy the applications on streaming data or large-scale datasets [19]. Practically, conducting FS can further enhance the performance, especially for large datasets [32]. Accordingly, there is a necessity to develop an algorithm with simpler mathematical computation, appropriately filters the heterogeneous features and also can scale to a large database. Therefore, developing scalable and effective FS algorithms is highly demanded to handle high-dimensional social media datasets. HPMR algorithm is suggested for solving these problems and can be applied efficiently for both supervised and unsupervised classification techniques.

3. PROPOSED ALGORITHM

HPMR mainly examined statistical evaluation and probability distribution methods as depicted in **Error! Reference source not found.**. However, while reducing feature space; there are two crucial requirements that must be considered: Firstly, the performance of classification technique must not be degraded substantially. Secondly, the task should be performed automatically since the manual procedure can be an extremely tedious effort. These issues are considered by applying the following algorithm.

Algorithm1:ALGO_High_Probability_Min_Redundancy
Input: D (M_0, M_1,……., M_{k-1})
            C (Pos, Neg)
            F(f_0,f_1,……., f_{n-1})

Output: Reduced features set (F_subset)
BEGIN
    H ← Apply Hapax legomenon (D)
    Construct a probabilistic model to find
        P ← Pos feature set (D, C)
        N ← Neg feature set (D, C)
Draw Venn diagram to find
    P U N
    P – N
    N – P
    P ∩ N
    for each f ∈ P ∩ N
        if (P(P) == P(N))
            ignore this feature.
                Else if (P(P) > P(N))
                return top Pos terms
                Else if (P(P) <P(N))
                return top Neg terms
        End if
    End for
  Extract outliers
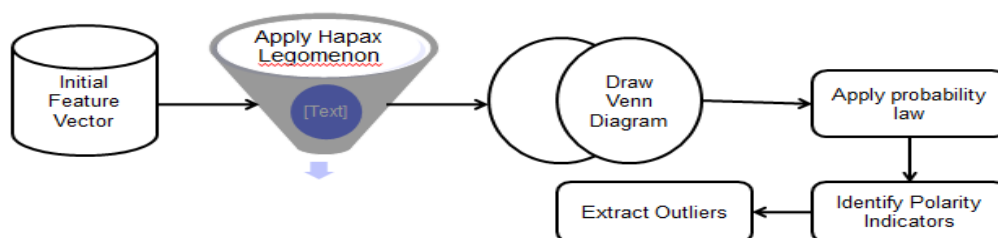 return F_{subset}
END

Fig 1: HPMR methodology

Practically, HPMR attempts to employ features that are relevant for prediction, but at the same time, it is important to have a set of features which is not redundant in order to increase robustness. Therefore, HPMR adjusts the relevance of the features by two major factors: first is minimal redundancy with selected features and the second is maximal relevance with the class. These factors are achieved by exploiting the subsequent procedure.

3.1 Hapax Legomenon

The algorithm starts by counting the number of times a word occurs overall the dataset and arranging these frequencies in a descending order. Actually, there is a large chunk of features occurs only once overall the feature space. In corpus linguistics, this phenomenon is known as a Hapax legomenon [9]. Because of users' different writing conventions, acronym usages, language differences, and spelling mistakes, there may be variations in the content of postings even if they are about the same event. As result, all rejected features have the lowest frequencies over all the feature space. Moreover, less common features are sparse and likely to cause over-fitting [1] [3]. The pruning heuristic gives also an indication to discard spelling mistakes, specific colloquial, informal text, irregular words, and lots of abbreviations that would be hard to categorize without any additional lexical vocabulary.

3.2 Venn Diagram

Not all frequent features are genuine or valuable for the mining process. There are also some uninteresting and redundant ones [27]. Redundancy pruning aims to remove these incorrect features. Venn diagrams presented in Fig 2 and set theory help identify different scenarios for the relationship between the candidate features f.
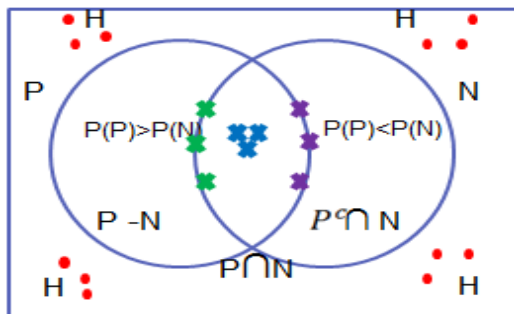

Fig 2: Venn diagram for feature space

Using set theory, the union of two sets P and N is the set of all features that belong to P or N (or both) and is denoted by P ∪ N.

$$P \cup N = \{x \mid x \in P \text{ or } x \in N\}, \qquad (1)$$

The intersection of two sets P and N is the set of all features that belong to both P and N and is denoted by P ∩ N.

$$P \cap N = \{x \mid x \in P \text{ and } x \in N\}, \qquad (2)$$

The difference of two sets P and N is the set of all features that belong to P and do not belong to N and is denoted by P - N.

$$P - N = \{x \mid x \in P \text{ and } x \notin N\}, \qquad (3)$$

The difference of two sets N and P is the set of all features that belong to N and do not belong to P and is denoted by N - P.

$$N - P = \{x \mid x \in N \text{ and } x \notin P\} \qquad (4)$$

Equation 1 expresses a new feature after applying Hapax Legonomon. The main assumptions were the following: (a) features are classified as positive (have probability 1 with positive instances), negative (have probability 1 with negative instances) and common (overlap in both categories with different probabilities) as established in Fig 3; (b) an heuristic function was exploited to pick out features, which controls the tradeoff between relevance and redundancy; (c) a greedy search strategy is used to achieve a balance between two desirable but incompatible features, and (d) the new feature subset is presumed optimal.

Obviously, variables should be selected according to their relevance. However, at the same time, redundancy should be avoided that appears when a feature essentially carries the same information for all classes. Eventually, equation 2 finds all features that overlap in both positive and negative instances with different probabilities. So, the following rules could be defined:

1. If the feature is distributed with equal probabilities along positive and negative instances (i.e. (P(P) = P(N)), then, it can be removed. This rule decreases the probability of mistaking important terms as redundant ones in the searching process. The function tries to capture the intuition that the best features for messages M_k are the ones distributed most in the sets of positive and negative M_k.
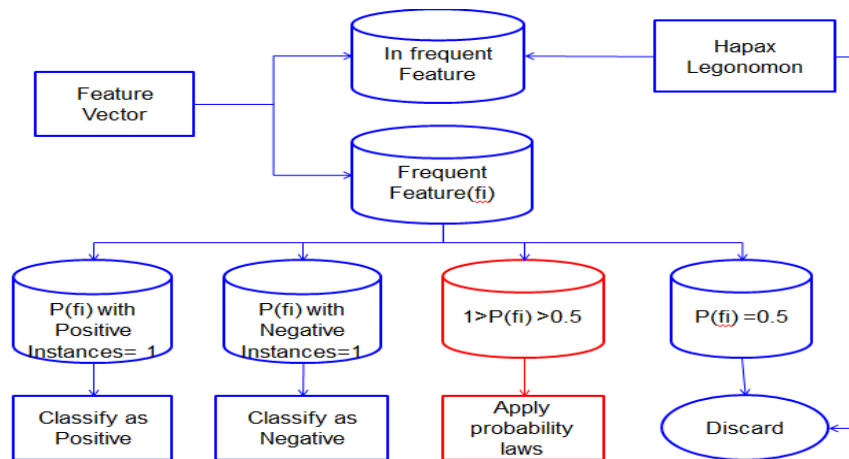

Fig 3: Applying probabilities laws

2. The attributes with a high probability of belongingness to dominantly one class are candidate to identify the category of the unknown instance. For example, if "جميل" word is occurring in 1000 positive instance and in 2 negative instances, then the value of this feature will be (1000-2)/(1000+2)= 0.996, its value is near to 1 indicates importance of the term in identifying the class of unknown messages. It indicates that if a new message is having "جميل" word, there is a high chance that this message belongs to a positive category.

3. On the other hand, all terms have almost equal probabilities of belongingness to both the categories are not useful in discriminating the class. Unfortunately, there are many words have low probabilities and almost equally distributed along each category (e.g. comes with positive instances with 51% and negative ones with 49%). such terms within some pre-defined range can be ignored.

The last two equations (3) and (4) are beneficial for the rule-based method, as they pick up the most frequently-occurring features in the corpus as polarity indicators. That is, if a feature is highly frequented and encountered in the positive instance with probability one, the unknown instances that have this feature could be tagged as positive. On the other hand, if a feature is highly frequented and appear only in the negative instance with probability one, the new instances would be elected to be negative. By applying these rules we can save a time consumed not only in the learning but also in the classification process.

3.3 Outliers

Moreover, some other few aspects which have already been detected as top views are called outliers. These terms appear frequently in several categories and are not specific for certain class. For the Arabic language, these words usually represent stop words that are not removed during preprocessing stage. Outliers can be checked and removed before classification according to some predefined threshold.

3.4 HPMR and social media SA

To emphasize the significance of HPMR, it has been relied upon to develop an intelligent system to perform SA for Arabic social media. The methodology starts by crawling raw data from different social media. The underlying dataset undergoes a simple filtering method that automatically excludes inconsistencies and out of scope comments. Then, morphological, semantic and emotion analysis are performed according to the study presented by [25]. A related task is the detection of subjectivity, a task that specifically separates factual information from opinionated ones.

Afterwards, the test set is passed through a series of preprocessing stages which applied prior to mining. It can extraordinarily improve the overall quality of the patterns and the required time for the actual mining [5]. Preprocessing techniques include tokenization, normalization, stemming, and stop word filtering [6]. The extracted feature sets have mainly adopted structural and social media-specific features. Emotion recognition [11], negation and punctuation were experimentally considered by the methodology proposed by [25].

3.5 Classification

Once a dataset has been reduced, the optimal feature vector needs to be combined with a classifier to check for accuracy. The strategy can be expressed by the following algorithm.

Algorithm 2: ALGO_SENTIMENT_ANALYSIS
Input: A trained dataset Dtrain
Output: Sentiment (Positive, Negative)
    BEGIN
      Read data from different social media
      Stripping out of scope messages
          for each message M_k do
              Decompose M_k to its message vector
              with i^{th} terms [M_{k1}, M_{k2}, .. M_{ki}]
              Extract features with some
              tokenization criteria and check for
              memory requirement
              Preprocessing
          End for
        Classify
       Compute the performance of the classifier
       Perform FS using HPMR and Chi test on D$_{train}$
      Evaluate the performance of each algorithm.
END

3.6 Merits of using HPMR

- Reliable for large-scale information processing.
- Has simple statistics evaluations and easier development and implementation.
- Reduces feature space up to 58%. Consequently, confidences memory requirements and CPU execution time.
- Can be applied for both supervised and unsupervised classification techniques.
- Capable of processing many thousands of features within minutes on a personal computer.
- Achieving a reasonable trade-off between relevance and redundancy
- The advantage HPMR algorithm over the remaining procedures is more remarkable for the massive sample sizes. This looks as a promising conclusion since huge sample sets are very common for web content classification, especially for social media.

4. CHI TEST

In addition to these qualitative merits, a quantitative quality indication might be obtained. Accordingly, the dataset is preprocessed using one of the conventional FS preprocessing filters (Chi Statistics). FS via Chi-square [24] is a very popular method for evaluating the independence between occurrences of two events [31]. Chi-squared attribute evaluates the worth of a feature by computing the value of the chi-squared statistic with respect to the class. In other words, it measures divergence from the distribution expected if one assumes the feature occurrence is actually independent of the class value. The initial hypothesis (Ho) is the assumption that the two features are unrelated, and it is tested by chi-square formula:

$$X^2 = \sum_{1=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})}{E_{ij}}$$

where r is the number of different values of the feature, c is the number of classes, Q_{ij} is the observed frequency with value i which are in class j, and E_{ij} is the expected frequency with value i and class j. In addition, the greater the value of X^2, the more unlikely it is that the distribution of values and classes are independent and the more relevant the feature is with respect to the class [8].

5. EXPERIMENT, RESULT, and EVALUATION

5.1 Dataset

The test set was crawled by searching Twitter API with specific queries including products' names, companies, and people. The original dataset consists of about 20000 messages. 13678 are considered as facts, news, duplicated and out of scope tweets. Actually, 6322 tweets are considered for training with the same number of positive and negative tweets. A balanced dataset is exploited to avoid the creation of models biased toward a specific class. The underlying tweets were tokenized into 14590 unique attributes as a baseline before applying any reduction methodology.

5.2 Comparison

The experiment is performed individually on three different stages. Firstly, the dataset is constructed into its original dimensionality (F_1) where a unigram feature set without any FS methods had taken as baseline accuracy. Secondly, the feature space is preprocessed using the FS preprocessing filter (Chi-square) to beget the new feature subset (F_2). Finally, the underlying feature vector is reduced by HPMR to generate the candidate feature subset (F_3). For each testbed, SVM is conducted as the classifier because of its reported performance [32] [14] as it gained practically a good track record for SA [4]. In all experiments, associated parameter settings and various kernels were investigated, and found that linear kernels yield the best performance. Then, the efficiency of all approaches had evaluated using standard classification metrics.

5.3 Evaluation

As expected, the highest frequent tokens (outliers) are the function words which occur in both positive and negative statements such as determiners, pronoun, prepositions, and conjunctions that impart very little meaning. The agglutinative nature of the Arabic language allows some specific determiner has to be more than 45 different morphological forms as indicated in Fig 4. Some of these forms might not be handled during preprocessing stage, so they can be removed manually from the new-born list before classification.

ان – اننا – انا – انه- انها انها انه –انهم – انهن-انهما- انك – انكم – انكما – انكن
فان- فاننا- فانا- فانه- فانها-فانهم- فانهن- فانهما- فانك –فانكم- فانكما – فانكن
لان- لاننا- لانا- لانه- لانها- لانهم- لانهن- لانهما- لانك –لانكم- لانكما – لانكن
وان- واننا- وانا- وانه – وانها – وانهم – وانهن - وانهما- وانك –وانكم- وانكما - وانكن

Fig 4: Example of outliers

However, other more informative tokens have high occurrence and represent the keywords of the underlying data set. Practically, frequent item identification allows the creation of two lists, 1) stop word list and 2) list of domain words. Domain words are dictionaries that collect lists of relevant named entities, important keywords, and phrases on certain topics. The selected terms are often hand-filtered for their relevance and importance. This function embodies the intuitions that (i) the more often a term occurs in a dataset, the more it is representative of its content, (ii) the more instances a term occurs in, the less discriminating it is. So, a fixed dictionary of the 100 most frequent tokens was created but ignores the 40 most frequent terms (outliers) from the original full vocabulary.

5.3.1 Accuracy measures

Table 1 highlights that performance of SVM on the considered dataset for all different stages. Initially, the feature vector was conducted to its original dimensionality and achieved 85.4% for F-Measure. Whenever a Chi-Square test is performed and combined with the best classifier, that is SVM method, the exploitation of the usual FS strategies slightly alters the quality of the results, instead of bringing up an added value. However, the sentiment classification accuracy drops by 1.2% compared to the baseline model trained on original feature space. Finally, after applying HPMR, its combination with SVM leads to excellent classification results. It demonstrates more robustness by providing suitable score F- Measure 89.9%.

Table 1 - Precision, Recall and F-Measure

Classifier	SVM		
Metrics	Precision%	Recall%	F-Measure%
F_1	85.3	85.6	85.4
F_2	84.3	84.4	84.2
F_3	89.9	89.9	89.9

As can be seen from the Table 1, the highest value of precision was achieved for F_3. The precision for the F_1 and F_2 was less than the F_3 by approximately 5 %. The worst recall value was obtained in the case of F_2. In words, using all features without applying any reduction methodology result in the classifier being too sensitive to the minor syntactical variations of the same feature and therefore these features are considered not correlated, which consequently affects the classifier performance.

5.3.2 Attribute reduction percentage

It is observed that Chi statistics successfully minimizes attributes to a very high extent reached about 70.9%, while HPMR compresses the initial data by approximately 58.6%. The summarized results are tabulated in **Error! Reference source not found.**2 and graphically represented in **Error! Reference source not found.**.

Table 2- Attribute reduction percentage

Features	# instances	# attributes	Attribute Reduction%
F_1		14590	0
F_2	6321	4251	70.9
F_3		6034	58.6

This reduction is a preprocessing step carried out prior to classification so that the classifiers can be designed in an easy way to compute. However while doing so, it must also retain the accuracy and must not lead to loss of information. It can greatly minimize memory issues and extensively affects training time.
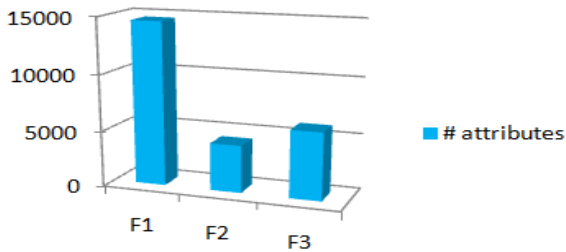
Fig 5: Reduction for all modes

5.3.3 Time consumption

Attribute weighting time is proportional to the number of attributes in a given dataset. The more features the higher the preprocessing time. Fig 6 illustrates the time taken for the mode formation in the three different stages. The lowest preprocessing time was achieved in the case of HPMR when compared with the other two vector types. That is, the simplicity of mathematical procedure delivers the diminished feature set in only a few second.

Obviously, F_1 consumes the highest computational time, and the resource consumption of F_1, in particular, was largely determined by the sample size. The training time sharply increased with larger sample sizes. F_1 requires approximately 10 hours, which is longer than the other methods. Chi statistics consumes training time for training more than the required time for HPMR.
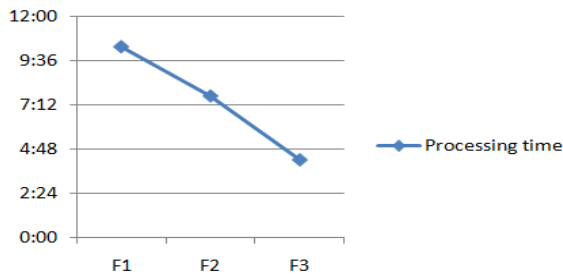

Fig 6: Processing time for all modes

6. CONCLUSION

Using of more effective FS would help improve the system's overall classification performance further especially for large scale social media contents. So, HPMR is presented as a new methodology for FS. The algorithm experimentally successes to redefine the feature space with reduction percentage reach 58% for Twitter and achieves about 90% for F-Measure. A quantitative comparison of HPMR approach and the classical Chi algorithm has been performed. Such comparison confirms the superiority of the HPMR. Practically, the quality of HPMR outperformed that of the Chi statistics from the viewpoint of accuracy measurements and CPU execution time. Additionally, it produces the best trade-off between accuracy and stability. Although Arabic language review data was used in this study, the proposed method can also be applied to other languages. Accordingly, such tool can aid the evaluation of customer opinions and measuring customers' loyalty about certain product, service, movie, or even political issue through the use of SA.

In the future, it would be valuable to further explore the potential of this approach on different domains and datasets. The algorithm should be investigated for other mining tasks, such as document articles, topic and genre classification. Additionally, there would be significant value in evaluating

more FS methods and combining some of those methods to examine if better feature sets can be produced.

7. REFERENCES

[1] Abbasi; H. Chen; A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums" ACM Trans. Information Systems, Vol. 26, No. 3, Article 12, June 2008. .

[2] Ahmed K. Farahat Ali Ghodsi Mohamed S. Kamel "An Efficient Greedy Method for Unsupervised Feature Selection" IEEE 11th International Conference on Data Mining. 2011.

[3] Andrea Esuli; Fabrizio Sebastiani; Ahmed Abasi,"AI and Opinion Mining, Part 2", IEEE Intelligent Systems. Vol. 25, No.4, Pp. 72 - 79, 2010.

[4] Anuj Sharma; Shubhamoy Dey "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis" Special Issue of International Journal of Computer Applications on Advanced Computing and Communication Technologies for HPC Applications - ACCTHPCA, pp. 15- 20, June 2012.

[5] Asha Gowda Karegowda, A. S. Manjunath & M.A.Jayaram "Comparative Study of Attribute Selection Using Gain Ratio and Correlation Based Feature Selection" International Journal of Information Technology and Knowledge Management, Vol. 2, No. 2, pp. 271-277, July-Dec. 2010.

[6] A. Alajmi; E. M. Saad; R. R. Darwish; "Toward an Arabic Stop-Words List Generation" International Journal of Computer Applications, Vol. 46, No.8, pp. 8 – 13, May 2012.

[7] A.M. Martinez; A.C. Kak, "PCA versus LDA," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, pp. 228-233, Feb. 2001.

[8] Bangsuk Jantawan, Cheng-Fa Tsai "A Comparison of Filter and Wrapper Approaches with Data Mining Techniques for Categorical Variables Selection" International Journal of Innovative Research in Computer and Communication Engineering, pp. 4501-4508, Vol. 2, No. 6, June 2014.

[9] Daniel Jurafsky, James H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. 2nd Edition. Prentice-Hall, 2009.

[10] Elhamifar, E.; Vidal, R. "Sparse Subspace Clustering: Algorithm, Theory and Applications" IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 35, No. 11, Pp. 2765 – 2781, 2013.

[11] Erik Cambria; "Affective Computing and Sentiment Analysis" IEEE Intelligent Systems. Vol. 31, No. 2, pp. 102 -107, Mar - April 2016.

[12] Fabrizio Sebastiani, "Machine Learning in Automated Text Categorization" ACM Computing Surveys, Vol. 34, No.1, pp. 1–47, Mar. 2002.

[13] Guangtao Wang; Qinbao Song; Baowen Xu; Yuming Zhou, "Selecting feature subset for high dimensional data via the propositional FOIL rules", Pattern Recognition Elsevier, Vol. 46, No. 1, pp. 199–214, Jan. 2013.

[14] G.Vinodhini; RM.Chandrasekaran "Sentiment Analysis and Opinion Mining: A Survey" International Journal of Advanced Research in Computer Science and Software Engineering. Vol. 2, No. 6, pp 282 - 292. June 2012.

[15] Huan Liu; Lei Yu "Toward integrating feature selection algorithms for classification and clustering" IEEE Trans. Knowledge and Data Engineering. Vol. 17, No. 4, Pp. 491 – 502, 2005.

[16] Jean-Charles Lamirel; Pascal Cuxac; Kafil Hajlaoui; Aneesh Sreevallabh Chivukula, "A new feature selection and feature contrasting approach based on quality metric: application to efficient classification of complex textual data" , PAKDD 2013 Int. Workshops on Trends and Applications in Knowledge Discovery and Data Mining. pp. 367–378, 2013.

[17] Jiguang Liang; Xiaofei Zhou; Li Guo1; Shuo Bai; "Feature Selection for Sentiment Classification Using Matrix Factorization", pp. 63 – 64. Florence, Italy, ACM. May 2015,

[18] Joseph D. Prusa; Taghi M. Khoshgoftaar; David J. Dittman "Impact of Feature Selection Techniques for Tweet Sentiment Classification" Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference. pp. 299 – 304. 2015.

[19] Jun Yan, Benyu Zhang, Shuicheng Yan , Qiang Yang, Hua Li; Zheng Chen; Wensi Xi; Weiguo Fan; Wei-Ying Ma; Qiansheng Cheng "IMMC: Incremental Maximum Margin Criterion" Aug., pp. 22-25, 2004, ACM, Aug. 2004.

[20] Jung-Yi Jiang; Ren-Jia Liou; Shie-Jue Lee, "A Fuzzy Self-Constructing Feature Clustering Algorithm for Text Classification" IEEE Trans. Knowledge and Data Engineering, Vol. 23, No. 3, Pp. 335 - 349 Jan. 2011.

[21] Junshi Xia; Lionel Bombrun; Tülay Adalı; Yannick Berthoumieu; Christian Germain; "Spectral Spatial Classification of Hyperspectral Images Using ICA and Edge- Preserving Filter via an Ensemble Strategy" IEEE Trans. Geoscience and Remote Sensing, Vol. 54, No. 8, pp. 4971 – 4982, 2016.

[22] J. Yan, B. Zhang, N. Liu, S. Yan, Q. Cheng, W. Fan, Q. Yang, W. Xi, and Z. Chen, "Effective and Efficient Dimensionality Reduction for Large-Scale and Streaming Data Preprocessing," IEEE Trans. Knowledge and Data Eng., Vol. 18, no. 2, pp. 1-14, Feb. 2006.

[23] K.Mugunthadevi, S.C. Punitha, M. Punithavalli "Survey on Feature Selection in Document Clustering

International Journal on Computer Science and Engineering (IJCSE)" Vol. 3 No. 3, Mar 2011.

[24] L.Ladha, T.Deepa "Feature Selection Methods and Algorithms" International Journal on Computer Science and Engineering (IJCSE) Vol. 3 No. 5, pp. 1787 – 1797, May 2011.

[25] Manal Mustafa; A.Shakour AlSamahi; Alaa Hamouda, "New Avenues in Arabic Sentiment Analysis", International Journal of Scientific & Engineering Research, Vol. 8, No. 2, pp. 907 - 915, Feb. 2017.

[26] Michael L. Raymer, William F. Punch, Erik D. Goodman, Leslie A. Kuhn, and Anil K. Jai "Dimensionality Reduction Using Genetic Algorithms" pp. 164 - 171 IEEE Trans. Evolutionary Computation, Vol. 4, No. 2, July 2000.

[27] M.A. Jawale; D.N. Kyatanavar; A.B. Pawar, "Automated Sentiment or Opinion Discovery System" International Journal of Computer Applications (IJCA), Vol. 106, No. 13, pp. 29- 35, Nov. 2014.

[28] Nicola Falco, Jon Atli Benediktsson, Lorenzo Bruzzone" Spectral and Spatial Classification of Hyperspectral Images Based on ICA and Reduced Morphological Attribute Profiles" IEEE Trans. Geoscience and Remote Sensing. Vol. 53, No. 11, pp. 6223 – 6240, 2015.

[29] Richard Jensen; Qiang Shen; "Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches", IEEE Trans. Knowledge and Data Engineering, Vol. 17, No. 1, pp. 1 – 15, Jan. 2005.

[30] Satoshi Niijima , Yasushi Okuno "Laplacian Linear Discriminant Analysis Approach to Unsupervised Feature Selection" IEEE/ACM Trans. Computational Biology and Bioinformatics, Vol. 6, No. 4, pp. 605-614 , Oct.-Dec. 2009.

[31] Tehseen Zia, Qaiser Abbas, Muhammad Pervez Akhtar " Evaluation of Feature Selection Approaches for Urdu Text Categorization" I.J. Intelligent Systems and Applications, pp. 33-40, June 2015.

[32] Yan Dang, Yulei Zhang, Hsinchun Chen "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews" IEEE Intelligent Systems, Vol. 25, No. 4, pp. 46- 53, July /Aug. 2010.

[33] Yijun Sun, Sinisa Todorovic, Steve Goodison, "Local Learning Based Feature Selection for High Dimensional Data Analysis", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, pp. 1610 – 1626. July 2010.