IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

Relevance Based Feature Selection Algorithm For Efficient Preprocessing of Textual Data Using HMM

R. Merlin Packiam[1*], V. Sinthu Janita Prakash[2]

[1,2]Department of Computer Science, Cauvery College for Women, *Tiruchirappalli*, Tamilnadu, India

[*]*Corresponding Author: rmerlinpackiam@gmail.com, Tel.: 9443673083*

Available online at: www.ijcseonline.org

Accepted: 19/Jan/2019, Published: 31/Jan/2019

Abstract— With a rapid growth of the world of Internet, the social media is eventually growing and is playing a very major role in most of our lives. There are various social networking sites such as Twitter, Google+, Face book which provide a platform for the people to present themselves. Twitter is an efficient micro-blogging tool which has become very popular throughout the world. Nowadays, there is an ongoing trend of posting every thought and emotion of one's life on these social networking sites. Due to this, emotion analysis has gained popularity in analyzing the thoughts, opinions, feelings, sentiments, etc., of various people. But handling such a huge amount of unstructured data is a tedious task to take up. Feature selection is the process of reducing the number of collected features to a relevant subset of features and is often used to combat the curse of dimensionality. This paper proposes a Relevance Feature Selection for efficient analytics on twitter data. After selecting the features from the tweets, Support Vector Machine (SVM) based classification is applied to analyze the data using Hidden Morkov Model(HMM). The performance of the proposed method has been evaluated through experiments. The entire research was evaluated through publicly available twitter data set with various metrics such as precision, recall, F-measure and Accuracy. By comparing the obtained results with the existing research results, the performance of the proposed work provides better result.

Keywords—Twitter, Bigdata,Feature Selection ,HMM

I. INTRODUCTION

Micro-blogging is, nowadays, one of the most well known specialized services among Internet clients. Every day, millions of users share opinions on various aspects of life. Micro-blogging users write about their life, share opinions on variety of topics and discuss current issues. [1][2]. Twitter has become one of the most popular micro-blogging services on the Internet. Users can send and read tweets which are small text based messages of up to 140 characters. The limited length of these tweet posts introduces some particularities regarding the text usage. For instance, abbreviations are often used, orthographic mistakes are made on purpose and hash-tags as well as emoticons are present in order to communicate the message of the author in a few words [3]. Twitter users are prone to spelling and typographical errors and to the use of abbreviations and slang. They may also use punctuation signs to emphasize their emotions like many exclamation marks. Usually, it is not necessary to include all terms of the initial form of a text in the machine learning step and some of them can be ignored, replaced, or merged with others. Thus, it arises the need of cleansing and normalizing the data, as their quality is a key factor to the success of the machine learning that follows pre-processing.

Due to the diverse nature of tweets, feature engineering methods for Twitter data can potentially generate tens of thousands of features, though each instance will only contain a few features of the entire feature set as tweets are limited to 140 characters in length. Feature selection techniques select a subset of features, much smaller than the total number of features, reducing computational time needed to train and classify tweets. Additionally feature selection can improve classifier performance by eliminating redundant or irrelevant features and reducing over fitting [4][5].

The ability of hidden Markov models (HMMs) to compensate for the variance in length of temporal sequences leads to good performance in speech processing, gesture recognition, DNA analysis, and other applications. An HMM is normally estimated by Maximum Likelihood Estimation (MLE). A HMM is composed of a sequence of unobserved states, modelled by a Markov chain, and a sequence of observed features related to the states. HMMs are generally used in applications where a sequence of unknown (hidden) states must be estimated from a sequence of correlated observations.

This paper proposes a novel procedure for feature subset selection. This approach is intended to work with hidden

Markov models (HMMs). HMMs have originally emerged in speech recognition. Recently they have become very popular in the entire domain. The main contribution of this work is to select the features from twitter data using HMM. In addition to this multiclass SVM is applied to verify the performance of selected features.

The rest of the paper is organized as follows. Section II explains the related work; Section III describes the proposed Reverence feature selection with classification; Section IV discusses the experimental results and Section V concludes the paper

II. RELATED WORK

Feature selection seeks to choose an optimal subset of features by eliminating features that are irrelevant or offer no additional information compared to features within the optimal subset.

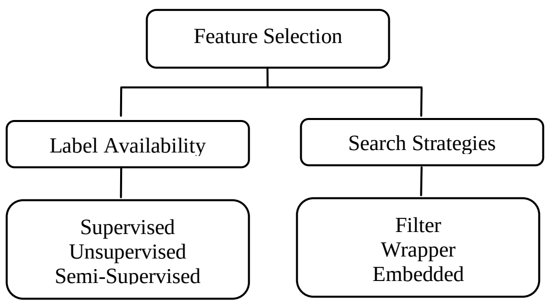

Figure 1 Feature Selection Categories

Figure 1 shows the categories of feature selection.

According to the availability of supervision (such as class labels in classification problems), feature selection can be generally classified as supervised (classification or regression), unsupervised (clustering) and semi-supervised methods [6]. Relating to different selection strategies, feature selection methods can be generally categorized as wrapper, filter and embedded methods.

This section explains some related work of feature selection.

Riham et al.,[7] select a compact feature subset from the exhaustive list of extracted features in order to reduce the computational complexity without sacrificing the classification accuracy. An online web-based feature selection tool (DWFS) was developed according to the GA (Genetic Algorithm) based wrapper paradigm [8]. DWFS also integrates various filtering methods that may be applied as a pre-processing step in the feature selection process. Wang et al., [9] developed a distance measure evaluating the difference between the selected feature space and all feature space to find a feature subset, which approximates all features. Nguyen et al., [10] proposes a new Gaussian based transformation rule for interpreting a particle as a feature subset, which is combined with the feature cluster based representation to develop a new PSO-based feature selection algorithm. A binary artificial bee colony (ABC) algorithm [11] for the feature selection problems, which is developed by integrating evolutionary based similarity search mechanisms into an existing binary ABC variant.

A Gini Index based feature selection method with Support Vector Machine (SVM) classifier is proposed for sentiment classification for large movie review dataset in [12]. Machine Learning Approach, was proposed in [13], which uses the bag-of-words (BoW) with the help of feature selection techniques which selects only important features by eliminating the noise and irrelevant features. In [14], the binary adaptation of cuckoo search known as the Binary Cuckoo Search is proposed for the optimum feature selection for a sentiment analysis of textual online content. Tommasel and Godoy [15] presented an online feature selection method for high-dimensional data based on the mixing of two information sources, social and content-based, for the real-time classification of short-text streams coming from social media.

Packiam et al.,[16] developed a Distance based feature selection algorithm(DFSA) for selecting minimum number of feature using diversity measure to enhance the accuracy. Packiam et al.,[17] proposed a optimum based feature selection algorithm(OFSA) for selecting optimum subset by the combination of Artificial Immune system(AIS) and Simulated Annealing(SA) algorithm to improve the classification accuracy.

To the best of our knowledge, there is no literature that specifically addresses feature selection for HMM in text mining field. But HMM is used in visual gesture recognition [18] and speech recognition [19]. Principal component analysis (PCA) can be used to reduce the number of features before the model parameters are estimated and compares PCA, discriminative feature analysis, sequential search and concludes that sequential search and discriminative feature analysis outperform PCA but suffer from higher computational load and require supervised data. Bashir et al., [20] introduces a novel method employing Gaussian mixture models (GMM)-based representations and hidden Markov model (HMM)-based classifiers for motion trajectory representation and analysis. Zhu *et al.* [21] use variational Bayesian (VB) methods to jointly estimate model parameters and select features for HMMs. This method does not require *a* priori knowledge of the number of states, or the number of mixtures if a Gaussian mixture model (GMM) is used for the emission distribution. This lack of information increases computation time, increases the complexity of the model, and, in some cases, can decrease parameter estimation accuracy. A vast majority of the techniques conduct feature selection during a pre-processing phase, which is independent of the model.

Roberto et al., [22] proposed a feature selection algorithm embedded in a hidden Markov model applied to gene expression time course data on either single or even multiple biological conditions. Adams et al., [23] propose a feature saliency hidden Markov model. Zheng et al [24]. Their strategy combines a hidden Markov model, a localized feature saliency measure and two t-Student distributions to describe the relevant and non-relevant features, to accurately model emission parameters for each hidden state.

In this proposed model relevant and irrelevant value calculated for each feature using Hidden Markov Model to select the relevant feature to increase the accuracy of the classifier.

III. METHODOLOGY

This research work proposes a Relevant based Feature Selection Algorithm (RFSA) for analyzing twitter data. First it pre-processes the collected tweets and apply HMM to select the relevant features(RFSA) and then SVM algorithm is used to classify the tweets. Figure 2 shows the proposed work flow.

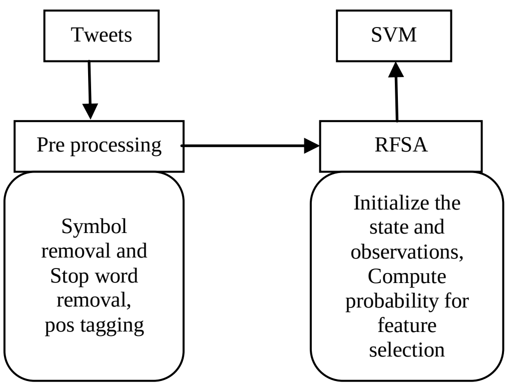

Figure 2 Proposed Approach Work Flow

The proposed method divided into following three steps.

- Preprocessing
- Relevance Feature Selection algorithm(RFSA)
- SVM Classification

Detail descriptions of the above steps are given in the following sub sections.

A. Tweets Preprocessing

Pre-processing is fundamental to all Natural Language Processing (NLP) Task. Steps needed for pre-processing of text in general depends on the targeted requirement or application. Here tweets are pre-processed for feature selection and analyzing the sentiments. Unlike other text documents, Twitter is a domain where people use their freedom to express the messages or comments in a flexible manner. Several attributes have been identified for a twitter status update or tweets. Maximum length of Twitter message is 140 characters which may include user-mention, hash tag, URL etc. The frequency of elongated words, slangs, acronyms and emoticons is much higher than any other domain.

Each tweet has an URL which represents the source of the content. Any user in tweets refers another user by providing @ symbol before the username. The tweets are associated with the name beginning with a # symbol. The tweets contains stop words like a, an, the, i, it, was etc., these entities are removed to improve the efficiency of the pre-processing. Then the stemming operation is performed to get the pure terms and tagging is performed. A set of terms has been selected as a observed feature set.

For Example, consider the tweets,

• Foursquare ups the game, just in time for #SXSW http://j.mp/grN7pK) - Still prefer @Gowalla by far, best looking Android app to date.
• Beautifully smart and simple idea RT @madebymany @thenextweb wrote about our #hollergram iPad app for #sxsw! http://bit.ly/ieaVOB
• I just noticed DST is coming this weekend. How many iPhone users will be an hour late at #SXSW come Sunday morning? #SXSW #iPhone

After preprocessing step,

• foursquare ups game just time still prefer far best looking android app date
• beautifully smart simple idea wrote ipad app
• just noticed dst coming weekend many iphone users will hour late come sunday morning

B. Relevance Feature Selection

Feature selection, as a dimensionality reduction technique, aims to choosing a small subset of the relevant features from the original ones by removing irrelevant, redundant or noisy features. Feature selection usually leads to better learning performance, i.e., higher learning accuracy, lower computational cost, and better model interpretability. This section describes a feature selection from twitter data set based on Hidden Markov Model.

The Hidden Markov Model is a finite set of *states*, each of which is associated with a (generally multidimensional) probability distribution. Transitions among the states are governed by a set of probabilities called *transition probabilities.* In a particular state an outcome or *observation* can be generated, according to the associated probability distribution.

Consider a HMM with continuous emissions and I states. Let $y=\{y_0,y_1,..y_T\}$ be the sequence of observed data, where each $y_t \in R^L$. The observation for the l-th feature at time t, which is represented by the l-th component of y_t is denoted by ylt. Let $x=\{x_0,x_1,\ldots,x_T\}$ be the unobserved state sequence The transition matrix of the Markov chain associated with this sequence is denoted as A. The components of this transition matrix are denoted by $a_{ij} = P(x_t=j|x_{t-1}=i)$, and π is used to denote the initial state distribution. In terms of these quantities, the complete data likelihood can be written as:

$$p(x, y \mid \Lambda) = \pi_{x_0} f_{x_0}(y_0) \prod_{t=1}^{T} a_{x_{t-1},x_t} f_{x_t}(y_t)$$

Where Λ is the set of model parameters, and $f_{x_t}(y_t)$ is the emission distribution of a given state x_t.

A feature is considered to be relevant if its distribution is dependent on the underlying state and irrelevant if its distribution is independent of the state. Let $r = \{r_1,r_2,\ldots, r_n\}$ be a set of binary variables indicating the relevancy of each feature. If $r_n = 1$, then the n-th feature is relevant. Otherwise, if $r_n = 0$, the n-th feature is irrelevant. Figure 2 shows the graphical representation of HMM for feature selection.

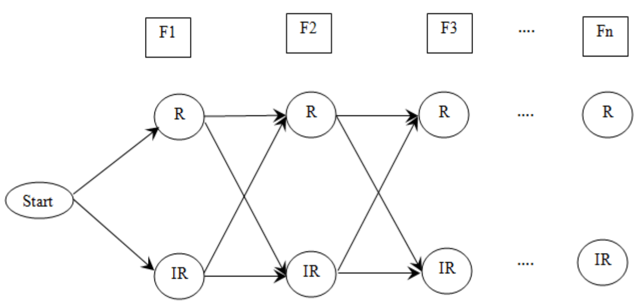

Figure 1 Graphical representation of HMM

Set two states as Relevant and Irrelevant. Set the Observation as extracted features after pre-processing step. The start probability of State as {'Relevant': 0.7, 'Irrelevant': 0.3}. The transition probability is initialized as,

T_{prob} = { ‘Relevant’: {'Relevant': 0.7, 'Irrelevant': 0.3},
‘Irrelevant’: {'Relevant': 0.4, 'Irrelevant': 0.6} }

The emission probability of each feature can be calculated as,

For Relevant State: $\dfrac{count(t \in D)}{count(D)}$ (1)

For Irrelevant State: $\dfrac{count(t \notin D)}{count(D)}$ (2)

Where D is the Total no of document and t is the current Feature

After calculating the emission probability, for each feature in the list compute probability for two states using equation (3) and (4).

For Relevant State:

PS.Prob * rel_ trans_p* rel_ emit_p (3)

For Irrelevant State:

PS.Prob * irrel_ trans_p* irrel_ emit_p (4)

Where PS.Prob = Previous State probability
rel_trans_p = relevant transition probability
irrel_trans_p = irrelevant transition probability
rel_emit_p = relevant emission probability
irrel_emit_p = irrelevant emission probability.

If the relevant state probability is high then the current feature is relevant otherwise not relevant.

The algorithm shows the pseudo code for RFSA using HMM

Algorithm: Relevance based Feature Selection Algorithm (RFSA)

Input: Features extracted after pre-processing step (F)
Output: Selected Relevant Features (SF)

1. Let *N* be the no of State (*S*) where *S(R) = Relevant* and *S(IR) = Irrelevant State*
2. Initialize Observation (*O*) as Features (*F*)
3. Let *T* be the no of observation *O*, where $O=\{f_1,f_2,\ldots,f_T\}$.
4. Let T_{prob} be the transition probability matrix with *i=1,2,..,N* rows and *j=1,2,..,N* columns
5. Let E_{prob} be the emission probability matrix with *i=1,2,..,N* rows and *j=1,2,..,T* columns
6. Assign state probability $S_{prob}(R)=0.7$ and $S_{prob}(IR)=0.3$
7. Assign value for T_{prob} where $\forall$ row in T_{prob}, $$\sum_{j=1}^{N} T_{prob_{i,j}} = 1$$
8. Calculate emission probability for state S(R) = $$E_{prob}(R,f_j) = \frac{count(f_j \in D)}{count(D)}$$
9. Calculate emission probability for state S(IR) = $$E_{prob}(IR,f_j) = \frac{count(f_j \notin D)}{count(D)}$$
10. $S(R)_{init}=S_{prob}(R) * E_{prob}(R,1)$;
11. $S(IR)_{init}=S_{prob}(IR) * E_{prob}(IR,1)$;
12. For each feature f_i in *F*
13. $R_{prob}=Max[S(R)_{init}*T_{prob}(R,R)*E_{prob}(R,f_i), S(R)_{init}*T_{prob}(R,IR)*E_{prob}(IR,f_i)]$
14. $IR_{prob} = Max[S(IR)_{init}*T_{prob}(IR,R)*E_{prob}(R,f_i), S(IR)_{init}*T_{prob}(IR,IR)*E_{prob}(IR,f_i)]$
15. Update the probability $S(R)_{init}=R_{prob}$ and $S(IR)_{init}=IR_{prob}$

16. If $R_{prob} > IR_{prob}$
    Rn=1
    Add Rn to SF
    Else
    Rn=0
17. End For

C. SVM Classification

After successful feature selection, the data set is classified using Multi-Class Support Vector Machine**.** Support Vector Machines (SVM) was originally designed for binary classification. But the twitter data comprises of more than two features, so multi class SVM classification is to need for twitter classification.

In general, the multi-class classification problems ($k > 2$) are commonly decomposed into a series of binary problems such that the standard SVM can be directly applied. Two representative ensemble schemes are one-versus-all (1VA) [25] and one-versus one (1V1) [26] approaches.

The one-versus-all (1VA) approach [25] constructs k separate binary classifiers for k-class classification. The m-th binary classifier is trained using the data from the m-th class as positive examples and the remaining $k-1$ classes as negative examples. During test, the class label is determined by the binary classifier that gives maximum output value. A major problem of the one-versus-rest approach is the imbalanced training set. Suppose that all classes have an equal size of training examples, the ratio of positive to negative examples in each individual classifier is $\frac{1}{k-1}$.

The one versus one (1V1) evaluates all possible pair-wise classifiers and thus induces $k(k-1)/2$ individual binary classifiers. Applying each classifier to a test example would give one vote to the winning class. A test example is labeled to the class with the most votes. The size of classifiers created by the one-versus-one approach is much larger than that of the one-versus-all approach. Compared with the one-versus-all approach, the one-versus-one method is more symmetric. HMM-SVM[27] approach combined the two techniques for the web news mining.

Here the text data is perfect suits for SVM classification due to the nature of the text. The number of classes defined in the SVM completely depends on the application, dataset used by the user. In this paper, SVM is used to classify the tweets based on the selected feature set.

IV. RESULTS AND DISCUSSION

This section explains the performance evaluation of proposed approach. The relevance based feature selection with SVM classification is implemented using R tool (version 3.5.1), and the experiments are performed on a Intel(R) Pentium machine with a speed 2.13 GHz and 2.0 GB RAM using Windows 7 32-bit Operating System.

For experiments, the real time twitter data set is extracted from https://data.world/datasets/twitter. The tweets are grouped into three emotion categories: Positive, Negative, No emotion. Table 1 shows the data set information

Table 1 Tweet Data Information

No of tweets	Positive	Negative	No emotions
9090	1573	242	7275

The proposed approach is evaluated using precision, recall, F-measure and accuracy.

Basic formula for the Precision ,Recall , F measure and Accuracy can be calculated as,

$$P = \text{Precision}= \frac{TP}{TP+FP}$$

$$R= \text{Recall}= \frac{TP}{TP+FN}$$

$$\text{F measure} = F = 2\times\frac{P.R}{P+R}$$

Accuracy = TP+TN/TP+FP+FN+TN

TP = True Positive, TN=True Negative,
FP= False Positive, FN= False Negative

The computational time is calculated for each stage of the entire framework such as pre-processing, feature selection and classification. The obtained result is shown in Table 2.

Table 2 Execution Time for Preprocessing, Feature Selection and Classification

Methods	Execution Time (Secs)
Preprocessing	2.454
Feature Selection	4.694
Classification	48.341

The evaluation metrics of precision, recall and f-measure shown in Table 3.

Table 3 Evaluation Metrics

Metrics	Value
Precision	0.91

Recall	0.99
F-Measure	0.94

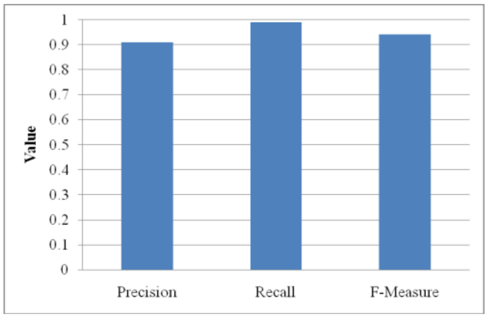

Figure 2 Precision, Recall, F-Measure

Table 4 Performance Evaluation of Classification Accuracy

Algorithm	Accuracy
DFSA	71.6
HMM-SVM	90.76
OFSA	98
RFSA	97.21

Table 4 shows the comparison of classification accuracy with previous approaches DFSA[16], OFSA[17], HMM-SVM[27] and the graphical representation shows in Figure 4

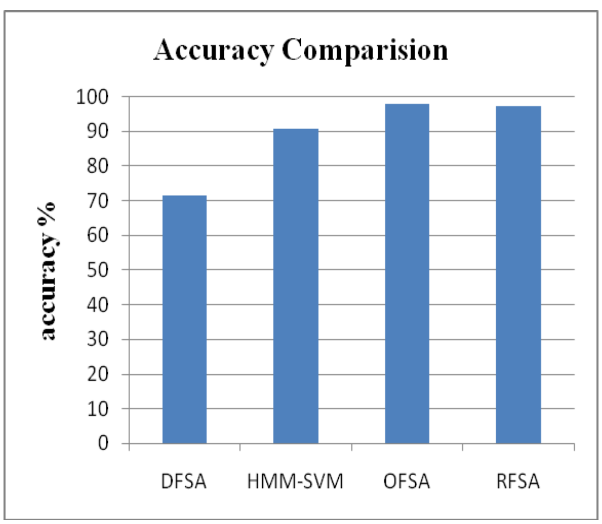

Figure 4 Accuracy Comparison

Overall Accuracy calculated as a sum of correct classification divided by the total number of classification. Precision , Recall , F1 score and Accuracy value calculated from confusion matrix with multiclass. The quality of the data increased by feature selection and Accuracy improved with previous work.

V. CONCLUSION AND FUTURE SCOPE

This paper presents a Relevance based Feature Selection Algorithm(RFSA) using Hidden Markov Model. This model first pre-processes the tweets and selects feature using RFSA. Based on these features, a SVM is applied to classify the tweets. Various feature selection algorithms and models were employed for classification. But the RFSA feature selection method was able to increase the accuracy of the classifier. The experiment use real time dataset for the performance from the results given in the tables, it is clear and noticed that the proposed framework is considered as an efficient approach for feature selection.

VI. REFERENCES

[1] Tsapatsoulis, Nicolas, and Constantinos Djouvas (2017), "Feature extraction for tweet classification: Do the humans perform better?.", Semantic and Social Media Adaptation and Personalization (SMAP), 2017 12th International Workshop on. IEEE, 2017.

[2] Packiam, R. Merlin, and V. Sinthu Janita Prakash. "An empirical study on text analytics in big data." Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on. IEEE, 2015.

[3] Anna Stavrianou, Caroline Brun, Tomi Silander, and Claude Roux, "NLP-based feature extraction for automated tweet classification". In Proceedings of DMNLP, Workshop at ECML/PKDD .7,2014.

[4] Prusa JD, Khoshgoftaar TM, Dittman DJ , "Impact of feature selection techniques for tweet sentiment classification", In: Proceedings of the 28th International FLAIRS Conference; 2015. p. 299–304,2015.

[5] harmendra Sharma, Suresh Jain, "Evaluation of Stemming and Stop Word Techniques on Text Classification Problem", International Journal of Scientific Research in Computer Science and Engineering Science and Engineering, Vol: 3, No. :2, 2015.

[6] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert Trevino, Jiliang Tang, and Huan Liu. "Feature selection: A data perspective". arXiv preprint arXiv:1601.07996, 2016

[7] Riham Mansour, Mohamed Farouk Abdel Hady,Eman Hosam, HaniAmr, and Ahmed Ashour , "Feature selection for twitter sentiment analysis: An experimental study", In International Conference on Intelligent Text Processing and Computational Linguistics, pages 92–103. Springer,2015.

[8] O. Soufan, D. Kleftogiannis, P.Kalnis, V. B. Bajic, and D.Gupta , "DWFS: a wrapper feature selection tool based on a parallel genetic algorithm", PLoSONE, vol. 10, no. 2, Article ID e0117988, 2015.

[9] S. Wang, W. Pedrycz, Q. Zhu, and W. Zhu, "Subspace learning for unsupervised feature selection via matrix factorization," Pattern Recognit. , vol. 48, no. 1, pp. 10–19, 2015.

[10] H.B.Nguyen, B.Xue, I.Liu, P. Andreae, and M. Zhang , "Gaussian transformation based representation in particle swarm optimisation for feature selection" in Applications of

Evolutionary Computation (LNCS 9028). Cham, Switzerland: Springer, pp. 541–553,2015.

[11] E. Hancer, B. Xue, D. Karaboga, and M. Zhang , "A binary ABC algorithm based on advanced similarity scheme for feature selection" Appl. Soft Comput., vol. 36, pp. 334–348, Nov. 2015

[12] Manek,A.S., Shenoy,P.D., Mohan,M.C., and Venugopal,K. , "Aspect term extraction for sentiment analysis in large movie reviews using Gini index feature selection method and SVM classifier", World Wide Web, 20 (2), 135–154,2017 .

[13] Agarwal B., Mittal N., "Machine Learning Approach for Sentiment Analysis", In: Prominent Feature Extraction for Sentiment Analysis. Socio-Affective Computing. Springer, Cham, 2016.

[14] Akshi Kumar, Shikhar Garg, Shobhit Verma and Siddhant Kumar(2019), "Sentiment Analysis Using Cuckoo Search for Optimized Feature Selection on Kaggle Tweets", International Journal of Information Retrieval Research (IJIRR) ,vol. 9, no. 1, 2019.

[15] A. Tommasel, D. Godoy "A Social-aware online short-text feature selection technique for social media " Inf. Fusion, 40 (pp. 1-17. 2018.

[16] Sinthu, R. Merlin Packiam, Dr V., and Janita Prakash. "Multilevel Sparse Dimension Selection Approach For Improved Big Data Processing Using Taxonomy." International Journal Of Innovation In Engineering Research And Management, ISSN 2348-4918, ISO 2000-9001 certified, E 4, no. 4 ,2017.

[17] Packiam, R. Merlin, and V. Sinthu Janita Prakash. "A Novel Integrated Framework Based on Modular Optimization for Efficient Analytics on Twitter Big Data." In Information and Communication Technology for Intelligent Systems, pp. 213-224. Springer, Singapore, 2019.

[18] J. A. V. Montero and L. E. S. Sucar , "Feature selection for visual gesture recognition using hidden Markov models" in Proc. 5th Int. Conf. Comput. Sci. (ENC), pp. 196-203, Sep. 2004..

[19] J. Nouza , "Feature selection methods for hidden Markov model-based speech recognition", in Proc. 13th Int. Conf. Pattern Recognit., vol. 2, pp. 186-190,1996.

[20] F. I. Bashir, A. A. Khokhar, and D. Schonfeld "Object trajectory-based activity classification and recognition using hidden Markov models", IEEE Trans. Image Process., vol. 16, no. 7, pp. 1912-1919, Jul. 2007

[21] H. Zhu, Z. He, and H. Leung , "Simultaneous feature and model selection for continuous hidden Markov models", IEEE Signal Process. Lett.,vol. 19, no. 5, pp. 279-282, May 2012.

[22] Roberto A. Cárdenas-Ovando, et al., "A feature selection strategy for gene expression time series experiments with hidden Markov models", bioRxiv preprint, 2018.

[23] Adams S, Beling P, Cogill R. Feature Selection for hidden Markov models and hidden Semi-Markov models. IEEE. Translations and content mining. Vol.4, Iss.1, pp. 1642–1657, Apr. 2016.

[24] Zheng Y, Jeon B, Sun L, Zhang J, Zhang H. Student's t-hidden Markov model for Unsupervised Learning Using Localized Feature Selection. IEEE Transactions on Circuits and Systems for Video Technology. Vol. 9,Iss.12, pp:1–10, July 2017

.[25] Vapnik, V. "Statistical Learning Theory". Wiley, New York (1998)

[26] Kreßel, U. "Pairwise classification and support vector machines". In: Schölkopf, B., Burges, C.,Smola, A. (eds.) Advances in Kernel Methods: Support Vector Learning, pp. 255–268. MIT Press, Cambridge 1999

[27] Krishnalal, G , Babu Rengarajan, S and G Srinivasagan, K . "A New Text Mining Approach Based on HMM-SVM for Web News Classification", International Journal of Computer Application. vol 1, Issue.9,2010.