ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

A survey on Attributes Selection for Sentiment Classification Using FRN and Dimension reduction models of MDD.

Sanjeev Kumar Singh, S. R. Yadav

P.G. Student, Computer Science & Engineering, Millennium Institute of Technology, Bhopal, MP, India. Assistant Professor& Head, Dept. ofC.S.E,Millennium Institute of Technology, Bhopal, MP, India.

ABSTRACT: The exploration of multidimensional data sets of all possible sizes and dimensions is a long-standing challenge in knowledge discovery from database, machine learning, and visualization. While multiple efficient visualization methods for n-D data analysis exist, the loss of information, occlusion, and clutter continue to be a challenge. A user-centric approach has been adopted in which user perception has been taken into consideration. Projection techniques has been focused that output 2D or 3D scatterplots that can then be used for a range of common data analysis tasks, which are categorized as pattern identification tasks, relation-seeking tasks, membership disambiguation tasks, or behaviour comparison tasks. The presence of noisy, irrelevant, and redundant attributes is major concern when incorporating large sets of diverse n-gram features for sentiment classification. These concerns can often make it difficult to harness the augmented discriminatory potential of extended feature sets. A rule-based multivariate text feature selection method called Feature Relation Network (FRN) has been proposed that considers semantic information and also leverages the syntactic relationships between n-gram features. FRN is intended to efficiently enable the inclusion of extended sets of heterogeneous n-gram features for enhanced sentiment classification. Experiments were conducted on three online review testbeds in comparison with methods used in prior sentiment classification research. FRN outperformed the comparison univariate, multivariate, and hybrid feature selection methods; it is able to select attributes resulting in significantly better classification accuracy irrespective of the feature subset sizes. Furthermore, by incorporating syntactic information about n-gram relations, FRN is able to select features in a more computationally efficient manner than many multivariate and hybrid techniques.

KEYWORDS: Multidimensional datasets, KDD, Scatterplots, FRN, Visualization, Segmentation and Sentiment Classification.

I. INTRODUCTION

Many procedures for n-D data analysis, knowledge discovery and visualization have demonstrated efficiency for different data sets [1–5]. However, the loss of information, occlusion, and clutter in visualizations of n-D data continues to be a challenge for knowledge discovery [1,2]. There is dimension scalability challenge for multidimensional data. Since only 2-D and 3-D data can be directly visualized in the physical 3-D world, visualization of n-D data becomes more difficult with higher dimensions as there is greater loss of information, occlusion and clutter, Further progress in data science will require greater involvement of end users in constructing machine learning models, along with more scalable, intuitive and efficient visual discovery methods and tools [6]. Visualization is a crucial step in the process of data analysis. Often, when analysing multidimensional data, dimensionality reduction (DR) techniques are displayed in form of 2D or 3D scatterplots that project the multidimensional points onto a lower-dimensional visual space.
The use of a rich set of n-gram features spanning many fixed and variable n-gram categories has been proposed. The extended feature set with a feature selection method has been coupled that is capable of efficiently identifying an enhanced subset of n-grams for opinion classification. Feature Relation Network proposed is a rule-based multivariate n-gram feature selection technique that efficiently removes redundant or less useful n-grams, allowing for more effective n-gram feature sets. FRN also incorporates semantic information derived from existing lexical resources, enabling augmented weighting/ranking of n-gram features. Experimental results reveal that the extended feature set and proposed feature selection method can improve opinion classification performance over existing selection methods.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

II. RELATED WORK

2D similarity-based layouts from a higher-dimensional space has been taken for many projection methods exist to generate it. The design goals include maintaining pairwise distances between points [6] as implemented in multidimensional scaling (MDS), maintaining distances within a cluster, or maintaining distances between clusters [7]. Principal component analysis (PCA) generates similarity layouts by reducing data to lower dimensional visual spaces [8]. Some projection methods, such as isometric feature mapping (Isomap), favour maintaining distances between clusters instead Isomap is an MDS approach that has been introduced as an alternative to classical scaling capable of handling non-linear data sets. It replaces the original distances by geodesic distances computed on a graph to obtain a globally optimal solution to the distance preservation problem [7]. Least-Square Projection (LSP) computes an approximation of the coordinates of a set of projected points based on the coordinates of some samples as control points. This subset of points is representative of the data distribution in the input space. LSP projects them to the target space with a precise MDS force-placement technique. It then builds a linear system from information given by the projected points and their neighbourhoods [9]. The correlations of data points or clusters are not always known after they have been mapped from a higher- dimensional data space to 2D or 3D display space. Thus, several approaches evaluate the best views of multidimensional data sets. Sips et al. [10] provide measures for ranking scatterplots with classified and unclassified data. They propose two additional quantitative measures on class consistency: one based on the distance to the cluster centroids, and another based on the entropies of the spatial distributions of classes. They propose class consistency as a measure for choosing good views of a class structure in high-dimensional space. Tan et al. [11], Paulovich et al. [9], and Geng et al. [12] also evaluate the quality of lay- outs numerically. By ranking the perceptual complexity of the scatterplots, other studies investigate user perception by conducting user studies on scatterplots, finding that certain arrangements were more pleasing to most users [13]. How- ever, these operational measures were not necessarily equivalent to the measures of user preference based on their qualitative perceptions. Sedlmair et al. [14] have discussed the influence of factors such as scale, point distance, shape, and position within and between clusters in qualitative evaluation of DR techniques. They examined over 800 plots in order to create a detailed taxonomy of factors to guide the design and the evaluation of cluster separation measures. They focused only on using scatterplot visualizations for cluster finding and verification.

DimStiller [15] is a system to provide global guidance for navigating a data-table space through the process of choosing DR and VE techniques. This analysis tool captures useful analysis patterns for analysts who must deal with messy data sets. Rensink and Baldridge [16] explore the use of simple properties such as brightness to generate a set of scatterplots in order to test whether observers could discriminate pairs using these properties. They found that perception of correlations in a scatterplot is rapid, and that in order to limit visual attention to specific information it is more effective to group features together. Etemadpour et al. [17] postulate that cluster properties such as density, shape, orientation, and size influence perception when interpreting distances in scatterplots, and specifically, observe that the density of clusters is more influential than their size. In general, little attention has been paid to providing details about low-level tasks that guide users to choose DR and VE techniques. However, both high-level goals and much more specific low-level tasks are important aspects of analytic activities. Amar et al. [18] presented a set of ten low-level analysis tasks that they found to be representative of questions that are needed to effectively facilitate analytic activity. Andrienko and Andrienko distinguish elementary tasks that address specific elements of a set and synoptic tasks thataddress entire sets or subsets, according to the level of analysis [19]. Brehmer and Munzer [20] emphasize three main questions, why the tasks are performed, how they are performed, and what are their inputs and output these questions encompass their concept of multi-level typology. They believe that low-level characterization does not describe the user's context or motivation; nor does it take into account prior experience and background knowledge." Their typology relies on a more abstract categorization based on concepts, rather than taxonomy of pre-existing objects or tasks. In contrast, we attempt to specify tasks at the lowest level that can provide details about multidimensional data projection. However, the general approach of Brehmer and Munzner can be easily adopted as a tool to put these low-level tasks in context, facilitating the evaluation of user experiences by evaluation designers. This approach provides essential information, such as motivation and user expertise, for field studies that examine visualization usage. Therefore, it has been shown that how the defined tasks can be described according to a typology of abstract tasks relating intents and techniques (how) to modes of goals and tasks (why). 1) Possible tasks performed has been categorized when analyzing a specific multidimensional data visualization, and 2) Guidelines for analysts to assist in selecting appropriate projection techniques for performing specific visualization tasks on data sets has been formulated.

Opinion mining involves several important tasks, including sentiment polarity and intensity assignment [18], Polarity assignment is concerned with determining whether a text has a positive, negative, or neutral semantic

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

orientation. Sentiment intensity assignment looks at whether the positive/negative sentiments are mild or strong. Given the two phrases "I don't like you" and "I hate you," both would be assigned a negative semantic orientation but the latter would be considered more intense. Effectively classifying sentiment polarities and intensities entails the use of classification methods applied to linguistic features. While several classification methods have been employed for opinion mining, Support Vector Machine (SVM) has outperformed various techniques including Naïve Bayes, Decision Trees, Winnow, etc. [21], [22], [27], [29]. The most popular class of features used for opinion mining is n-grams [28], [38]. Various n-gram categories have attained state-of-the-art results [23], [27]. Larger n-gram feature sets require the use of feature selection methods to extract appropriate attribute subsets. Next, these two areas have been discussed: n-gram features and feature selection techniques used for sentiment analysis.

III. METHODOLOGY

A. N-Gram Features for Sentiment Analysis

N-gram features can be classified into two categories: fixed and variable. Fixed n-grams are exact sequences occurring at either the character or token level. Variable n-grams are extraction patterns capable of representing more sophisticated linguistic phenomena. A plethora of fixed and variable n-grams have been used for opinion mining, including word, part-of-speech (POS), character, legomena, syntactic, and semantic n-grams. Word n-grams include bag-of-words (BOWs) and higher order word n-grams (e.g., bigrams, trigrams). Word n-grams have been used effectively in several studies [28]. Typically, unigrams to trigrams are used [23], [27], though 4-grams have also been employed [34]. Word n-grams often provide a feature set foundation, with additional feature categories added to them [24], [27], [34], [38]. Given the pervasiveness of adjectives and adverbs in opinion-rich text, POS tag, n-grams are very useful for sentiment classification [30], [32]. Additionally, some studies have employed word plus part-of-speech (POS Word) n-grams. These n-grams consider a word along with its POS tag in order to overcome word-sense disambiguation in situations where a word may otherwise have several senses [38]. For example, the phrase "quality of the" can be represented with the POS Word trigram "quality-noun of- prep the-det." Character n-grams are letter sequences. For example, the word "like" can be represented with the following two and three letter sequences "li, ik, ke, lik, ike." While character n-grams were previously used mostly for style classification, they have recently been shown to be useful in related affect classification research attempting to identify emotions in text [22]. Legomena n-grams are collocations that replace once (hapax legomena) and twice occurring words (dis legume -na) with "HAPAX" and "DIS" tags [22], [38].

TABLE 1

N-Gram Features Used for Sentiment Analysis

N-Gram Category	Examples	Prior Studies
Character	q, u, qu, ua, al, li, qua, ual, ali	[2]
Word	quality, quality of, quality of the	[1, 8, 25, 28, 38, 27]
POS Tag	noun, noun prep, noun prep det	[12, 28]
Word/POS Tag	quality-noun of-prep the-det	[38]
Legomena	the UNIQUE, of the UNIQUE different-adj U-noun	[2] [37, 38]
Syntactic Phrase Patterns	<sub> passive-verb DECL::NP VERB NP <subj>ActInfVP	[33][12][34]
Semantic Phrase Patterns	SYN125 ofthe strong-tyranny, weak-berrationn+dj, av+n POSITIVE of thePP/Appreciation:ORI/Negative	[6][33][10][27][4]

Hence, the trigram "I hate Jim" would be replaced with "I hate HAPAX" provided "Jim" only occurs once in the corpus. The intuition behind such collocations is to remove sparsely occurring words with tags that will allow the extracted n-grams to be more generalizable [37], [38].
Syntactic phrase patterns are learned variable n-grams [34]. Ril off et al. [33] developed a set of syntactic templates and information extraction patterns (i.e., instantiations of those templates) reflective of subjective content. Given a set of predefined templates, patterns with the greatest occurrence difference across sentiment classes are extracted. For

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

example, the template “<subj> passive-verb” may produce the pattern “<subj> was satisfied.” Such phrase patterns can represent syntactic phenomena difficult to capture using fixed-word n-grams [32], [38].
Semantic phrase patterns typically use an initial set of terms or phrases, which are manually or automatically filtered and coded sentiment polarity/intensity information. Many studies have used WordNet to automatically generate semantic lexicons [39], [43] or semantic word classes [46]. Riloff et al. [53] used a semiautomated approach to construct sets of strong/weak subjectivity and objective nouns. Others have manually annotated or derived semantic phrases [54], [50].
Table 1 provides a summary of n-gram features used for opinion classification. Based on the table, we can see that many n-gram categories have been used in prior opinion mining research. However, few studies have employed large sets of heterogeneous n-grams. As stated before, most studies utilized word n-grams in combination with one other category, such as POS tag, legomena, semantic, or syntactic n-grams, e.g., [41], [44], [47], [54], [58].

B. **Feature Selection for Sentiment Analysis**

Prior sentiment classification studies have placed limited emphasis on feature selection techniques, despite their benefits [40]. Feature selection can potentially improve classification accuracy [37], narrow in on a key feature subset of sentiment discriminators, and provide greater insight into important class attributes. There are two categories of feature selection methods [35], [36], both of which have been used in prior sentiment analysis work: univariate and multivariate.Univariate methods consider attributes individually. Examples include information gain, chi-squared, log likelihood, and occurrence frequency [41]. Although univariate methods are computationally efficient, evaluating individual attributes can also be disadvantageous since important attribute interactions are not considered. It is also easier to interpret the contribution of individual attributes using univariate methods. Most opinion mining studies have used univariate feature selection methods such as minimum frequency thresholds and the log-likelihood ratio [32], [47], [59]. Information gain (IG) [64], [65] has also been shown to work well for various text categorization tasks, including sentiment analysis [43]. Tsutsumi et al. [55] used the Chi Squared test to select features for text sentiment classification. Table 2 shows select univariate feature selection methods used in sentiment classification studies.

TABLE 2

1. **Univariate Methods Used for Sentiment Classification**

Chi Squared [35]
$\chi^2(a,Y) = \sum_{a_{x_j}\in\{0,1\}} \sum_{i\in Y} \frac{\left(F(a_{x_j}, Y=i) - E(a_{x_j}, Y=i)\right)^2}{E(a_{x_j}, Y=i)}$ where : $\chi^2(a,Y)$ is the chi - squared value for feature a across classes Y $X = [x_1, x_2, \ldots x_m]$ are the training examples $a_{x_j} = 1$ if the training instance x_j contains feature a, $a_{x_j} = 0$ otherwise $F(a_{x_j}, Y=i)$ is the observed frequency of a_x, when $Y=i$ $E(a_{x_j}, Y=i) = \frac{p(a)p(Y=i)}{m}$ is the expected value of a_x, when $Y=i$, across X
Information Gain [2, 3, 13]
$IG(Y,a) = H(Y) - H(Y \mid a)$ where : $IG(Y,a)$ is the information gain for feature a $H(Y) = -\sum_{i\in Y} p(Y=i)\log_2 p(Y=i)$ is the entropy across classes Y $H(Y \mid a) = -\sum_{j\in a} p(a=j) \sum_{i\in Y} p(Y=i \mid a=j)\log_2 p(Y=i \mid a=j)$ is the entropy of $Y \mid a$
Log Likelihood Ratio [12, 27, 39]
$w(a) = \max_i \left(p(a \mid Y=i) \log \frac{p(a \mid Y=i)}{p(a \mid \neg Y=i)} \right)$ where : $w(a)$ is the log likelihood for feature a across classes Y

Multivariate methods consider attribute groups or subsets. These techniques sometimes use a wrapper model for attribute selection, where the accuracy of a target classifier is used as an evaluation metric for the predictive power of a particular feature subset [36]. Examples include decision tree models, recursive feature elimination, and genetic algorithms. By performing group-level evaluation, multivariate methods consider attribute interactions.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

Consequently, these techniques are also computationally expensive in relation to univariate methods. Decision tree models (DTMs) use a wrapper, where a DTM is built on the training data and features incorporated by the tree are included in the feature set [41]. Recursive feature elimination uses a wrapper model based on an SVM classifier [35]. During each iteration, the remaining features are ranked based on the absolute values of their SVM weights, and a certain number/percentage of these are retained [42], [43], [44]. Genetic algorithms (GAs) have been used to search for ideal subsetsacross the feature subspace in text classification problems such as style [40] and sentiment analysis [43]. A major pitfall associated with GA is that they can be computationally very expensive, since hundreds/thousands of solutions have to be evaluated using a classifier [43]. Feature subsumption hierarchies (FSHs) use the idea of performance-based feature subsumption to remove redundant or irrelevant higher order n-grams [54]. Only those word bigrams and trigrams are retained, which provide additional information over the unigrams they encompass. Table 3 shows multivariate methods used for sentiment classification.

TABLE 3

2. Multivariate Methods Used for Sentiment Classification

Decision Tree Models [1]
Given training examples $X = [x_1, x_2, \ldots x_m]$ and class labels $y = [y_1, y_2, \ldots y_m]$
Initialize subset of surviving features $s = [1, 2, \ldots p]$ and ranked feature list $r = []$
Repeat until $s = []$ or some stopping criterion has been reached
Train the classifier $\alpha = DT(X, y)$
Extract the decision tree features from α into $d = [s(a_1), s(a_2), \ldots s(a_t)]$
Update the ranked feature list $r = [r, d]$
Update the surviving features $s = s(1: a_1 - 1, a_1 + 1: a_2 - 1, a_2 + 1: a_t - 1, a_t + 1: p)$
Output ranked feature list r
Feature Subsumption Hierarchy [34]
Initialize subset of surviving features $s = [1,2,\ldots p]$
Repeat until all potential feature subsumptio ns have been evaluated
If feature s(a) representationally subsumes feature s(b) based on the hierarchy
If $IG(Y, s(a)) \geq IG(Y, s(a)) + \delta$
Eliminate s(b) from the feature set $s = s(1: b-1, b+1: p)$
where : δ is a parameter
$IG(Y, s(a))$ is the informatio n gain for feature $s(a)$ across classes Y
Genetic Algorithm [3]
Given training examples $X = [x_1, x_2, \ldots x_m]$ and class labels $y = [y_1, y_2, \ldots y_m]$
Initialize solution population $s = [s_1, s_2, \ldots s_r]$, where $s_x = [s_{x1}, s_{x2}, \ldots s_{xp}]$, $s_{xn} \in \{0,1\}$
and the feature s_{xn} is included in solution s_x if $s_{xn} = 1$
Repeat until some stopping criterion has been reached
Initialize population of new solutions $t = [t_1, t_2, \ldots t_r]$, where $t_x = []$
Evaluate each solution's fitness $F_{s_x} = Fitness(s_x, X, y)$
Select solutions based on fitness and add to t, where $p(s_x \in t) \propto F_{s_x} (\sum_{i=1}^{r} F_{s_i})^{-1}$
For each of the $r/2$ solution pairs in t
If random number $q \in [0,1] < thresh_c$, crossover t_x and t_{x+1} at point k
$t_x = [t_x(t_{x1} : t_{xk}), t_{x+1}(t_{x+1k+1} : t_{x+1p})]$ and $t_{x+1} = [t_{x+1}(t_{x+11} : t_{x+1k}), t_x(t_{xk+1} : t_{xp})]$
For each of the r solutions in t
For each t_{xn} in t_x, if $u \in [0,1] < thresh_m$, mutate t_{xn} where $t_{xn} = 1 - t_{xn}$
Set s equal to t for the next iteration $s = t$
Recursive Feature Elimination [3]
Given training examples $X = [x_1, x_2, \ldots x_m]$ and class labels $y = [y_1, y_2, \ldots y_m]$
Initialize subset of surviving features $s = [1, 2, \ldots p]$ and ranked feature list $r = []$
Repeat until $s = []$
Train the classifier $\alpha = SVM(X, y)$
Compute the weight vector $w = \sum_k \alpha_k y_k x_k$
Compute the ranking criteria $c_i = \lvert w_i \rvert$, for all i
Find the feature with the smallest ranking criterion $a = \arg\min(c)$
Update the ranked feature list $r = [s(a), r]$
Eliminate the feature with the smallest ranking criterion $s = s(1: a-1, a+1: p)$
Output ranked feature list r

3.Other Feature Selection Methods

In addition to prior sentiment feature selection methods, it is important to briefly discuss multivariate and hybrid methods used in related tasks. Principal component analysis (PCA) has been used considerably for dimensionality reduction in text style classification problems [66]. Recently, many powerful dimensionality reduction techniques have

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

also been applied to non-text feature selection problems. These include conditional mutual information (CMIM), harmonic mean, geometric mean, general averaged divergence analysis, and discriminative locality alignment (DLA) [14], [67], [68], [69], [70]. CMIM outperformed comparison techniques (including DTM) on image classification and biomedical prediction tasks [14]. DLA outperformed methods such as PCA and linear discriminant analysis on image classification tasks [70].Hybrid methods that combine univariate measures with multivariate selection strategies can potentially improve the accuracy and convergence efficiency of otherwise slower multivariate methods [45], [42]. For instance, a hybrid GA utilizing the IG measure has been shown to converge faster than regular GA, when applied to feature sets spanning up to 26,000 features [43].

IV.RESEARCH GAPS

Based on the review, appropriate gaps have been identified. Most studies have used limited sets of n-gram features, typically employing one or two categories [47], [48]. Larger n-gram feature sets introduce computational difficulties and potential performance degradation stemming from noisy feature sets. For instance, the popular 2,000 movie review testbed developed by Pang et al. 48] has over 49,000 bag-of-words [44]. Higher order n-gram feature spaces can be even larger, with hundreds of thousands of potential attributes. Feature selection methods are needed to help manage the large feature spaces created from the use of heterogeneous n-grams. As Riloff et al. [54] noted, using additional text features without appropriate selection mechanisms is analogous to “throwing the kitchen sink.” However, large-scale feature selection requires addressing relevanceand redundancy, something many existing methods.
Redundancy is a big problem since there are a finitenumber of attributes that can be incorporated and n-grams tend to be highly redundant by nature. In thecase of univariate methods, redundant features occupy valuable spots that may otherwise be utilized by attributes providing additional information and discriminatory potential. Powerful multivariatemethods are capable of alleviating redundancy; however, they are often unsuitable for computational reasons. These methods have typically been applied to smaller feature sets, e.g., [35], [40]. It is unclear whether hybrid feature selection methods have thepotential to overcome issues stemming from redundancy. More-very, most of the feature electionmethods described are generic techniques thathave been applied to a plethora of problems, sincethey assess attribute relevance solely based on the training data. Whenever possible, domain knowledge should be incorporated into the feature selection process [36]. Existing lexicons and knowledge bases pertaining to the semantic and syntactic properties of n-grams could be exploited for enhanced assessment of relevance and redundancy associated with text attributes.

TABLE 4

A. N-Gram Feature Set

Label	Description	Examples	
N-Char	Character- level n-grams	1-Char	I, L, 0, V, E, C, H, 0, C, 0, L, A
		2-Char	LO, OV, VE, CH, HO, OC, CO, OL
		3-Char	LOV, OVE, CHO, HOC, OCO
N-Word	Word-level n-grams	1-Word	I, LOVE, CHOCOLATE
		2-Word	I LOVE, LOVE CHOCOLATE
		3-Word	I LOVE CHOCOLATE
N-POS	Part-of-speech tag n-grams	1-POS	I, ADMIRE_VBP, NN
		2-POS	ADMIRE_VBP NN
		3-POS	I ADMIRE VBP NN
N-POSWord	Word and POS tag n-grams	1-POSWord	LOVE ADMIRE_VBP
		2-POSWord	I I LOVE ADMIRE_VBP
		3-POSWord	I LOVE ADMIRE_VBP CHOCOLATE NN
N-Legomena	Hapax legomena and Dis legomena n-grams	2-Legomena	LOVE DIS
		3-Legomena	I LOVE DIS
N-Semantic	Semantic class n-grams	1-Semantic	SYN-Pronoun, SYN-Affection
		2-Semantic	SYN-Pronoun SYN-Affection
		3-Semantic	SYN-Pronoun SYN-Affection SYN-Candy

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

IEP-A/E	Information extraction patterns	IEP-A	<possessive> NP, <subj>AuxVP AdjP, <subj>AuxVPDobj, ActVP<dobj>, ActVP Prep <np>
		IEP-B	<subj>PassVP, InfVP Prep <np>, InfVP<dobj>
		IEP-C	<subj>ActVP
		IEP-D	<subj>ActVPDobj
		IEP-E	<subj>ActInfVP, <subj>PasslnfVP, ActInfVP<dobj>

1. ExtendedN-Gram Feature Set: A rich set of n-gram features has been incorporated, consisted of all the categories discussed in the literature review. The feature set is shown in Table 4. The syntactic n-grams were derived using the Sundance package [53], [54]. This tool extracts n-gram instantiations of predefined pattern templates. Sundance learns n-grams that have the greatest occurrence difference across user-defined classes. For instance, the n-gram "endorsed <dobj>" is generated from the pattern template "ActVP<dobj>." The semantic n-grams were derived using WordNet, following an approach similar to that used by Kim and Hovey [39] and Mishne [43]. Words are clustered into semantic categories based on the number of common items in their sunsets. New words are added to the cluster with the highest percentage of synonyms in common provided the percentage is above a certain threshold. Otherwise, the word is added to a new cluster.

B. Feature Relation Network

For text n-grams, the relationship between n-gram categories can facilitate enhanced feature selection by considering relevance and redundancy, two factors critical to large-scale feature selection [61]. We propose a rule-based multivariate text feature selection method that considers semantic information and also leverages the syntactic relationships between n-gram features in order to efficiently remove redundant and irrelevant ones. Comparing all features within a feature set directly with one another can be an arduous endeavour. However, if the relationship between features can be utilized, thereby comparing only some logical subset of attributes, then the feature selection process can be made more efficient. Given large quantities of heterogeneous n-gram features, the FRN utilizes two important n-gram relations: Subsumption and parallel relations. These two relations enable intelligent comparison between features in a manner that facilitates enhanced removal of redundant and/or irrelevant n-grams

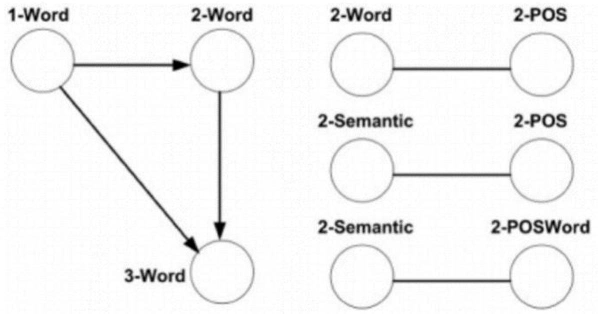


Fig 1 (left) Subsumption relations between word n-gram and (right) parallel relations between various bigrams.

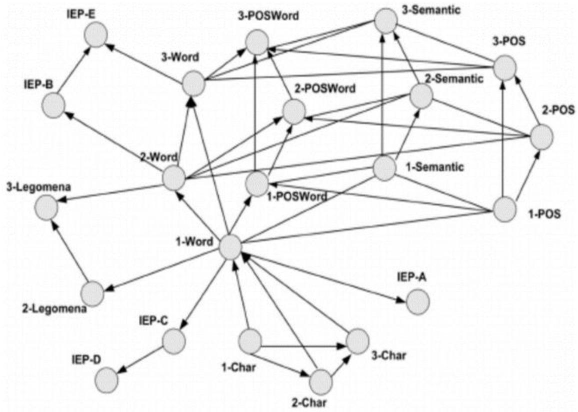


Fig. 2. The feature relation network

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

1. **Subsumption Relations**

In addition to prior sentiment feature selection methods, it is important to briefly discuss ultivariate. The notion of subsumption was originally proposed by Riloff et al. [34]. A subsumption relation occurs between two n-gram feature categories, where one category is a more general, lower order form of the other [34]. A subsumes BðA!BÞif B is a higher order n-gram category whose n-grams contain the lower order n-grams found in A. For example, word unigrams subsume word bigrams and trigrams, while word bigrams subsume word trigrams (as shown on the left side of Fig. 1). Given the sentence "I love chocolate," there are six-word n-grams: I, LOVE, CHOCOLATE, I LOVE, LOVE CHOCOLATE, and I LOVE CHOCOLATE. The unigram LOVE is obviously important, generally conveying positive sentiment. However, what about the bigrams and trigrams? It depends on their weight, as defined by some heuristic (e.g.,log likelihood or information gain). We only wish to keep higher order n-grams if they are adding additional information greater than that conveyed by the unigram LOVE. Hence, given A!B, we keep features from category B if their weight exceeds that of their general lower order counterparts found in A by some threshold t [34]. For instance, the bigrams I LOVE and LOVE CHOCOLATE would only be retained if their weight exceeded that of the unigram LOVE by t(i.e., if they provided additional information over the more general unigram). Similarly, the trigram I LOVE CHOCOLATE would only be retained if its weight exceeded that of the unigram LOVE and any remaining bigrams (e.g., I LOVE and OVE CHOCOLATE) by t.

2. Parallel Relations

A parallel relation occurs where two heterogeneous same order n-gram feature groups may have some features with similar occurrences. For example, word unigrams (1-Word) can be associated with many POS tags (1-POS), and vice versa. However, certain word and POS tags' occurrences may be highly correlated. Similarly, some POS tags and semantic class unigrams may be correlated if they are used to represent the same words. For example, the POS tag ADMIRE_VP and the semantic class SYN-Affection both represent words such as "like" and "love." Given two n-gram feature groups with potentially correlated attributes, A is considered to be parallel to B (A—B). If two features from these categories A and B, respectively, have a correlation coefficient greater than some threshold p, one of the attributes is removed to avoid redundancy. The right side of Fig. 1 shows some examples of bigram categories with parallel relations.

Correlation is a commonly used method for feature selection [31], [37]. However, correlation is generally used as a univariate method by comparing the occurrences of an attribute with the class labels, across instances [31]. Comparing attribute intercorrelation could remove redundancy, yet is computationally infeasible, often necessitating the use of search heuristics [37], [40]. FRN allows the incorporation of correlation information by only comparing select n-grams (ones from parallel relation categories within the FRN).

3. The Complete Network

Fig. 2 shows the entire FRN, consisted of the nodes previously described in Table 3. The network encompasses 22 n-gram feature category nodes and numerous subsumption and parallel relations between these nodes. The detailed list of relations is presented in Table 5. The order in which the relations are applied is important to ensure that redundant and irrelevant attributes are removed correctly. Subsumption relations are applied prior to parallel relations. Furthermore, subsumption relations between n-gram groups within a feature category are applied prior to across category relations (i.e., 1-Word !2-Word isappliedpriorto1-Word !1-POSWord).

TABLE 5

List of Relations between N-Gram Feature Groups

Feature Group	Relations
Subsumption Relations	
N-Char	1-Char 4 2-Char, 1-Char 4 3-Char, 2-Char 4 3-Char
N-Word	1-Word 4 2-Word, 1-Word 4 3-Word, 2-Word 4 3-Word
N-POS	1-POS 4 2-POS, 1-POS 4 3-POS, 2-POS 4 3-POS
N-POSWord	1-POSWord 4 2-POSWord, 1-POSWord 4 3-POSWord, 2-POSWord 4 3-POSWord
N-Legomena	2-Legomena 4 3-Legomena
N-Semantic	1-Semantic 4 2-Semantic, 1-Semantic 4 3-Semantic 2-Semantic 4 3-Semantic

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

IEP-A/E	1-Word 4 IEP-A, 1-Word 4 IEP-C, IEP-C 4 IEP-D, 2-Word 4 IEP-B, 3-Word 4 IEP-E, IEP-B 4 IEP-E
Char-Word	1-Char 4 1-Word, 2-Char 4 1-Word, 3-Char 4 1-Word
Word-POSWord	1-Word 4 1-POSWord, 2-Word 4 2-POSWord, 3-Word 4 3-POSWord
POS-POSWord	1-POS 4 1-POSWord, 2-POS 4 2-POSWord, 3-POS 4 3-POSWord
Word-Legomena	1-Word 4 2-Legomena, 2-Word 4 3-Legomena
Parallel Relations	
Word-POS	1-Word — 1-POS, 2-Word — 2-POS, 3-Word — 3-POS
Word-Semantic	1-Word — 1-Semantic, 2-Word — 2-Semantic, 3-Word — 3-Semantic
POS-Semantic	1-POS — 1-Semantic, 2-POS — 2-Semantic, 3-POS — 3-Semantic
POSWord-Semantic	1-POSWord — 1-Semantic, 2-POSWord — 2-Semantic, 3-POSWord — 3-Semantic

4. **Feature Weights: Incorporating Semantic Information**

Features weights w(ax) computed by considering their occurrence distribution across classes in the training datawt(ax), as well as their semantic weight ws(ax), which is based on the degree of subjectivity associated with the n-gram. Utilizing the semantic weight in addition to the training weight is intended to enhance relevance measurement and alleviate over fitting attributable to solely relying on training data for the calculation of feature weights.An n-gram's potential level of subjectivity is derived from SentiWordNet, a lexical resource that contains three sentiment polarity scores (i.e., positivity, negativity, and objectivity) for synsets consisted of word-sense pairs [29]. entiWordNet contains scores for over 150,000 words, with scores being on a 0-1 scale. For instance, the synset consisting of the verb form of the word "short" and the word "short-change" has a positive score of 0 and a negative score of 0.75.The semantic weight ws(ax) for an n-gram is computed by Determining the average polarity value across the individual tokens encompassed within the n-gram. For each token ax the polarity values are the average of the sum of its positive and negative scores for each word-sense pair in receiving a semantic weight fig 4 describes the FRN algorithm details. Given feature a from category A; we first find the feature categories that are subsumed by A. Then, all features from these categories containing

Let $A = \{a_1, a_2, \ldots a_n\}$ denote a set of word n - grams //e.g., 1 - Word or 2 - Word

For each a_x, where $a_x = (a_{x1}, \ldots a_{xd})$ denotes a tuple in A, the weight for a_x is :

$$w(a_x) = wt(a_x) + ws(a_x)$$

Where $wt(a_x)$ is the weight for feature a_x in the training data,

given v and w are part of the set of c class labels, $v \neq w$, and $c \geq 2$:

$$wt(a_x) = \max_{v,w}\left(P(a_x \mid v) \log\left(\frac{P(a_x \mid v)}{P(a_x \mid w)} \right) \right)$$

And $ws(a_x)$ is the semantic weight for feature a_x :

$$ws(a_x) = \frac{1}{d} \sum_{i=1}^{d} \left(\frac{1}{k} \sum_{j=1}^{k} s(a_{xi}, j) \right)$$

Where $s(a_{xi}, j)$ is the sum of the positive and negative scores

for the word a_{xi} and j is one of the k senses of a_{xi} in Senti WordNet

Fig. 3. Weighting Mechanism for n-grams.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

Let $A = \{a_1, a_2, \ldots a_n\}$ and $B = \{b_1, b_2, \ldots b_m\}$ denote two sets of n - grams
//e.g., 1 - Word, 3 - Word, etc.

if $A \rightarrow B$ //A subsumes B
 For each a_x, where $a_x = (a_{x1}, \ldots a_{xd})$ denotes a tuple in A with $w(a_x) > 0$
 Let $C \subseteq B$, where $C = \{c_1, c_2, \ldots c_y\}$
 And each $c_x = (c_{x1}, \ldots c_{xe})$ denotes a tuple in C with $w(c_x) > 0$
 Where the tuple a_x is a part of each c_x
  if $s(a_x) = s(c_x)$ //check the semantic orientation of the two features
   if $w(a_x) \geq w(c_x) - t$
    $w(c_x) = 0$

if $A - B$ //A is parallel to B
 For each a_x, where $a_x = (a_{x1}, \ldots a_{xd})$ denotes a tuple in A with $w(a_x) > 0$
 Let $C \subseteq B$, where $C = \{c_1, c_2, \ldots c_y\}$
 And $c_x = (c_{x1}, \ldots c_{xe})$ denotes a tuple in C with $w(c_x) > 0$
 Where each c_x is potentially correlated with a_x
  if $\mathrm{Corr}(a_x, c_x) \geq p$
   if $w(a_x) \geq w(c_x)$ then $w(c_x) = 0$
   if $w(a_x) < w(c_x)$ then $w(a_x) = 0$

Where :
$\mathrm{Corr}(a, b)$ is the correlation coefficient for a and b across the m training instances :

$$\mathrm{Corr}(a,b) = \frac{\sum_{x=1}^{m}(a_x - \bar{a})(b_x - \bar{b})}{\sqrt{\sum_{x=1}^{m}(a_x - \bar{a})^2 \sum_{x=1}^{m}(b_x - \bar{b})^2}}$$

$w(a_x)$ is the weight for feature a_x, computed as described in Fig. 3

$$s(a_x) = \arg\max_{v,w}\left(P(a_x \mid v)\log\left(\frac{P(a_x \mid v)}{P(a_x \mid w)}\right)\right)$$

t and p are predefined thresholds //we used $t = 0.05$ and $p = 0.90$

Fig 4. FRN Algorithm

Orientation	Sentence
Positive	I LOVE THIS DIGITAL CAMERA
Positive	I LOVE THE POWERFUL LENS A LOT TOO
Positive	I REALLY LIKE COMPACT SIZE OF THIS CAMERA
Negative	I DON’T LIKE THE AUTOFOCUS FUNCTION TONIGHT
Negative	THE BATTERY LIFE ALSO LEAVES MUCH TO BE DESORED
Negative	THE FLIMSY CAMERA DESIGN IS JUST NOT VERY FLATTERNG

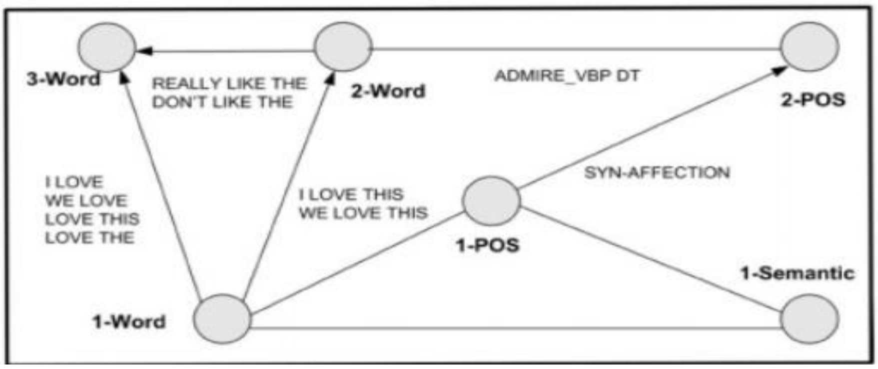
(a)

Feature	Category	Orientation	Weight	FRN Weight
LOVE	1-Word	Positive	1.0000	1.0000
I LOVE	2-Word	Positive	1.0000	0
WE LOVE	2-Word	Positive	1.0000	0
LOVE THIS	2-Word	Positive	1.0000	0
LOVE THE	2-Word	Positive	1.0000	0
I LOVE THIS	3-Word	Positive	1.0000	0

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

WE LOVE THE	3-Word	Positive	1.0000	0
REALLY LIKE	2-Word	Positive	1.0000	1.0000
DON'T LIKE	2-Word	Negative	1.0000	1.0000
REALLY LIKE THE	3-Word	Positive	1.0000	0
DON'T LIKE THE	3-Word	Negative	1.0000	0
ADMRE_VBP	1-POS	Positive	1.0000	0.8239
SYN_AFFECTION	1-Sentence	Positive	1.0000	0
ADMRE_VBP DT	2-POS	Positive	1.0000	0
LIKE THE	2-Word	None	0.8239	0
LIKE	1-Word	None	0.8239	0

(b)



(c)

TABLE 6

Description of Online Review Testbeds

Test Bed	Source	#Reviews	#Classes
Digital Cameras	www.epinions.com	2.000	5(1-5 stars)
Automobiles	www.edmunds.com	2.000	5(1,3,5,7,9 Stars)
Movies	www.rottentomatoes.com	2.000	2(Positive,Negative)

the substring a and having the same semantic orientation are retrieved. The semantic orientation of a feature is defined as the class for which the attribute has the highest probability of occurring. The semantic orientation of features is compared to avoid having features such as DON'T LIKE get subsumed by the unigram LIKE (since the two features have opposing semantic orientations). Feature weights are computed using the procedure described in the prior section and Fig. 3. The weights for the retrieved features are compared against that of a, and only those features are retained with a weight greater than aby some threshold t.

The parallel relations are enforced as follows: Given feature a from category A, we find the feature categories that are parallel to A. Features from these categories with potential co-occurrence with a are retrieved. The correlation coefficient for these features is computed in comparison with a. If the coefficient is greater than or equal to some threshold p, one of the features is removed. We remove the feature with the lower weight (ties are broken rbitrarily). It is important to note that for subsumption and parallel relations, only features still remaining in the feature set are analyzed and/or retrieved (i.e., ones with a weight greater than 0).

Although FRN utilizes subsumption relations as does FSH, it differs from FSH [34] in many ways. First, FRN incorporates seven n-gram feature categories whereas FSH only employs word n-grams and information extraction patterns. Second, FSH utilizes a weighting function that incorporates a unique training data-based weighting heuristic wtðaxÞ and a semantic weighting heuristic based on an independent lexicon wsðaxÞ, while FSH utilizes the feature'sIG score. Third, FRN incorporates subsumption and parallel relations, while FSH only uses subsumption. Fourth, FRN represents relations in a network, where features from any category can potentially be removed.In contrast, FSH uses a tree representation, where all features from the highest-level node (i.e., word unigrams) are always retained.

Fig. 5 shows an illustration of the FRN applied to a six-sentence testbed (three positive and three negatively oriented sentences). The table in the bottom left corner shows the feature weights for many key categories (e.g.,word, POS, and

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 5, Issue 5 , May 2018

semantic n-grams). The weights depicted include the initial wðaxÞ, the wtðaxÞ based on the six-sentence testbed, wsðaxÞ, and the adjusted wðaxÞ after the FRN has accounted for redundancy. The FRN is able to remove redundant or less useful n-grams, keeping only 6 of the 16 features shown. For example, the bigram I LOVE gets subsumed by the unigram LOVE. Similarly, the semantic class unigram SYN-Affection is parallel to the POS tag ADMIRE_VBP, and therefore, removed. Details for each removed n-gram are provided in the FRN on the right-hand side of the diagram. It is important to note that only the portion of the FRN, which is relevant to these features, is shown. The removed n-grams are placed next to the subsumption or parallel relation responsible for their removal. These features correspond to the features with an adjusted wðaxÞ of 0

V. CONCLUSION

Visualization will always remain a combination of art and science. While quantitative techniques can guide interpretation, there is still a need for managerial insight in order to make business decisions from visualizations In this study, the use of FRN for improved selection of text attributes has been proposed for enhanced sentiment classification. FRN's use of syntactic relation and semantic information regarding n-grams enabled it to achieve improved results over various univariate, multivariate, and hybrid feature selection methods.

REFERENCES

[1]. Bertini, E.; Tatu, A.; Keim, D. Quality metrics in high-dimensional data visualization: An overview and systematization. IEEE Trans. Vis. Comput. Gr. 2011, 17, 2203–2212.
[2]. Ward, M.; Grinstein, G.; Keim, D. Interactive Data Visualization: Foundations, Techniques, and Applications; A K Peters/CRC Press: Natick, MA, USA, 2010.
[3]. Rubel, O.; Ahern, S.; Bethel, E.W.; Biggin, M.D.; Childs, H.; Cormier-Michel, E.; DePace, A.; Eisen, M.B.; Fowlkes, C.C.; Geddes, C.G.; et al. Coupling visualization and data analysis for knowledge discovery from multi-dimensional scientific data. Procedia Comput. Sci. 2010, 1, 1757–1764.
[4]. Inselberg, A. Parallel Coordinates: Visual Multidimensional Geometry and Its Applications; Springer: New York, NY, USA, 2009.
[5]. Wong, P.; Bergeron, R. 30 years of multidimensional multivariate visualization. In Scientific Visualization—Overviews, Methodologies and Techniques; Nielson, G.M., Hagan, H., Muller, H., Eds.; IEEE Computer Society Press: Washington, DC, USA, 1997; pp. 3–33.
[6]. Kovalerchuk, B.; Kovalerchuk, M. Toward virtual data scientist. In Proceedings of the 2017 International Joint Conference on Neural Networks, Anchorage, AK, USA, 14–19 May 2017; pp. 3073–3080.
[7]. Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction.
[8].Joline, I. T. (1986). Principal Component Analysis. Springer-Verlag.
[9].Paulovich, F. V., Nonato, L. G., Minghim, R., and Levkowitz, H. (2008). Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. IEEE Transactions on Visualization and Computer Graphics, 14(3):564{575.
[10].Sips, M., Neubert, B., Lewis, J. P., and Hanrahan, P. (2009). Selecting good views of high-dimensional data using class consistency. Computer Graphics Forum (Proc. EuroVis 2009), 28(3):831{838.
[11].Tan, P.-N., Steinbach, M., and Kumar, V. (2005). Introduction to Data Mining. Addison-Wesley Longman, Boston, MA, USA.
[12].Geng, X., Zhan, D.-C., and Zhou, Z.-H. (2005). Supervised nonlinear dimensionality reduction for visualization and classi_cation. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 35(6):1098{1107.
[13].Tatu, A., Bak, P., Bertini, E., Keim, D. A., and Schneidewind, J. (2010). Visual quality metrics and human perception: an initial study on 2D projections of large multidimensional data. In Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '10), pages 49-56.
[14].Sedlmair, M., Brehmer, M., Ingram, S., and Munzner, T. (2012a). Dimensionality reduction in the wild: Gaps and guidance - ubc computer science technical report tr-2012-03. Technical report, The University of British Columbia.
[15].Ingram, S., Munzner, T., Irvine, V., Tory, M., Bergner, S., and Mller, T. (2010). Dimstiller:Workows for dimensional analysis and reduction. In IEEE VAST, pages 3-10. IEEE.
[16].Rensink, R. A. and Baldridge, G. (2010). The perception of correlation in scatterplots. Comput. Graph. Forum, 29(3):1203-1210.
[17].Etemadpour, R., Olk, B., and Linsen, L. (2014d). Eye-tracking investigation during visual analysis of projected multidimensional data with 2d scatterplots. In 5th International Conference on Information Visualization Theory and Applications (IVAPP), pages 233-246, Lisbon, Portugal
[18]. Amar, R., Eagan, J., and Stasko, J. (2005). Low-level components of analytic activity in information visualization. In Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization, INFOVIS '05, pages 15-, Washington, DC, USA. IEEE Computer Society.
[19]. Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., and Wrobel, S. (2011). A conceptual framework and taxonomy of techniques for analyzing movement. J. Vis. Lang. Comput., 22(3):213-232.
[20]. Brehmer, M. and Munzner, T. (2013). A multi-level typology of abstract visualization tasks. IEEE Trans. Visualization and Computer Graphics (TVCG) (Proc. InfoVis), 19(12):2376-2385.
[21] A. Abbasi and H. Chen, "CyberGate: A System and Design Framework for Text Analysis of Computer Mediated Communication," MIS Quarterly, vol. 32, no. 4, pp. 811-837, 2008.
[22] A. Abbasi, H. Chen, S. Thoms, and T. Fu, "Affect Analysis of Web Forums and Blogs Using Correlation Ensembles," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 9, pp. 1168-1180, Sept. 2008.
[23] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," ACM Trans. Information Systems, vol. 26, no. 3, article no. 12, 2008.
[24] S. Argamon, C. Whitelaw, P. Chase, S.R. Hota, N. Garg, and S. Levitan, "Stylistic Text Classification Using Functional Lexical Features," J. Am. Soc. Information Science and Technology, vol. 58, no. 6, pp. 802-822, 2008.
[25] P.V. Balakrishnan, R. Gupta, and V.S. Jacobs, "Development of Hybrid Genetic Algorithms for Product Line Designs," IEEE Trans. Systems, Man, and Cybernetics, vol. 34, no. 1, pp. 468-483, Feb. 2004.

[26] A. Burgun and O. Bodenreider, "Comparing Terms, Concepts, and Semantic Classes in WordNet and the Unified Medical Language System," Proc. North Am. Assoc. Computational Linguistics Workshop, pp. 77-82, 2001.
[27] H. Cui, V. Mittal, and M. Datar, "Comparative Experiments on Sentiment Classification for Online Product Reviews," Proc. 21st AAAI Conf. Artificial Intelligence, pp. 1265-1270, 2006.
[28] S.R. Das and M.Y. Chen, "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," Management Science, vol. 53, no. 9, pp. 1375-1388, 2007.
[29] A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining," Proc. Fifth Conf. Language Resources and Evaluation, pp. 417-422, 2006.
[30] Z. Fei, J. Liu, and G. Wu, "Sentiment Classification Using Phrase Patterns," Proc. Fourth IEEE Int'l Conf. Computer Information Technology, pp. 1147-1152, 2004.
[31] G. Forman, "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," J. Machine Learning Research, vol. 3, pp. 1289-1305, 2004.
[32] M. Gamon, "Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis," Proc. 20th Int'l Conf. Computational Linguistics, pp. 841- 847, 2004.
[33] M. Genereux and M. Santini, "Exploring the Use of Linguistic Features in Sentiment Analysis," Proc. Corpus Linguistics Conf., pp. 27-30, 2007.
[34] F. Fleuret, "Fast Binary Feature Selection with Conditional Mutual Information," J. Machine Learning Research, vol. 5, pp. 1531-1555, 2004.
[35] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification Using Support Vector Machines," Machine Learning, vol. 46, pp. 389-422, 2002.
[36] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," J. Machine Learning Research, vol. 3, pp. 1157- 1182, 2003.
[37] M. Hall and L.A. Smith, "Feature Subset Selection: A Correlation Based Filter Approach," Proc. Fourth Int'l Conf. Neural Information Processing and Intelligent Information Systems, pp. 855-858, 1997.
[38] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," Proc. ACM SIGKDD, pp. 168-177, 2004.
[39] S. Kim and E. Hovy, "Determining the Sentiment of Opinions," Proc. 20th Int'l Conf. Computational Linguistics, pp. 1367-1373, 2004.
[40] J. Li, R. Zheng, and H. Chen, "From Fingerprint to Writeprint," Comm. ACM, vol. 49, no. 4, pp. 76-82, 2006.
[41] H. Liu and H. Motada, Feature Extraction, Construction, and Selection—Data Mining Perspective. Kluwer Academic Publishers, 1998.
[42] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 4, pp. 491-502, Apr. 2005.
[43] G. Mishne, "Experiments with Mood Classification," Proc. Stylistic Analysis of Text for Information Access Workshop, 2005.
[44] D. Mladenic, J. Brank, M. Grobelnik, and N. Milic-Frayling, "Feature Selection Using Linear Classifier Weights: Interaction with Classification Models," Proc. ACM SIGIR, pp. 234-241, 2004.
[45] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining Product Reputations on the Web," Proc. ACM SIGKDD, pp. 341-349, 2002.
[46] T. Nasukawa and T. Nagano, "Text Analysis and Knowledge Mining System," IBM Systems J., vol. 40, no. 4, pp. 967-984, 2001.
[47] V. Ng, S. Dasgupta, and S.M.N. Arifin, "Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews," Proc. Conf. Computational Linguistics, Assoc. for Computational Linguistics, pp. 611-618, 2006.
[48] B. Pang, L. Lee, and S. Vaithyanathain, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 79-86, 2002.
[49] B. Pang and L. Lee, "A Sentimental Education: Sentimental Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc.42nd Ann.MeetingoftheAssoc.Computational
Linguistics, pp. 271-278, 2004.
[50] F. Peng, D. Schuurmans, V. Keselj, and S. Wang, "Automated Authorship Attribution with Character Level Language Models," Proc. 10th Conf. European Chapter of the Assoc. Computational Linguistics, 2003.
[51] A. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews," Proc. Human Language Technology, Empirical Methods in Natural Language Processing, pp. 339-346, 2005.
[52] E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 105-112, 2003.
[53] E. Riloff, J. Wiebe, and T. Wilson, "Learning Subjective Nouns Using Extraction Pattern Bootstrapping," Proc. Seventh Conf. Natural Language Learning, pp. 25-32, 2003.
[54] E. Riloff, S. Patwardhan, and J. Wiebe, "Feature Subsumption for Opinion Analysis," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 440-448, 2006.
[55] K. Tsutsumi, K. Shimada, and T. Endo, "Movie Review Classification Based on Multiple Classifier," Proc. 21st Pacific Asia Conf. Language, Information, and Computation, pp. 481-488, 2007.
[56] P.D. Turney and M.L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association," ACM Trans. Information Systems, vol. 21, no. 4, pp. 315-346, 2003
[57] J. Wiebe, T. Wilson, and M. Bell, "Identifying Collocations for Recognizing Opinions," Proc. Assoc. for Computational Linguistics, European Chapter of the Assoc. for Computational Linguistics Workshop Collocation, 2001.
[58] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, "Learning Subjective Language," Computational Linguistics, vol. 30, no. 3, pp. 277-308, 2004.
[59] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment Analyzer: Extracting Sentiments about a Given Topic Using Natural Language Processing Techniques," Proc. Third IEEE Int'l Conf. Data Mining, pp. 427-434, 2003.
[60] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," Proc. 20th Int'l Conf. Machine Learning, pp. 856-863, 2003.
[61] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," J. Machine Learning Research, vol. 5, pp. 1205-1224, 2004.
[62] Z. Zhang, "Weighing Stars: Aggregating Online Product Reviews for Intelligent E-Commerce Applications," IEEE Intelligent Systems, vol. 23, no. 5, pp. 42-49, Sept. 2008.
[63] I.H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, second ed. Morgan Kaufmann, 2005.
[64] C.E. Shannon, "A Mathematical Theory of Communication," Bell Systems Technical J., vol. 27, no. 10, pp. 379-423, 1948.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 5, Issue 5 , May 2018

[65] J.R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.
[66] A. Abbasi and H. Chen, "Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace," ACM Trans. Information Systems, vol. 26, no. 2, article no. 7, 2008.
[67] W. Bian and D. Tao, "Harmonic Mean for Subspace Selection," Proc. 19th Int'l Conf. Pattern Recognition, 2008.
[68] D. Tao, X. Li, X. Wu, and S.J. Maybank, "Geometric Mean for Subspace Selection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 260-274, Feb. 2009.
[69] D. Tao, X. Li, X. Wu, and S.J. Maybank, "General Averaged Divergence Analysis," Proc. Seventh IEEE Int'l Conf. Data Mining, pp. 302-311, 2007.
[70] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch Alignment for Dimensionality Reduction," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 9, pp. 1299-1313, Sept. 2009