CrossMark

Hybrid sentiment classification on twitter aspect-based sentiment analysis

Nurulhuda Zainuddin[1] · Ali Selamat[1,2,3] · Roliana Ibrahim[1]

Abstract Social media sites and applications, including Facebook, YouTube, Twitter and blogs, have become major social media attractions today. The huge amount of information from this medium has become an attractive resource for organisations to monitor the opinions of users, and therefore, it is receiving a lot of attention in the field of sentiment analysis. Early work on sentiment analysis approached this problem at a document-level, where the overall sentiment was identified, rather than the details of the sentiment. This research took into account the use of an aspect-based sentiment analysis on Twitter in order to perform a finer-grained analysis. A new hybrid sentiment classification for Twitter is proposed by embedding a feature selection method. A comparison of the accuracy of the classification by the principal component analysis (PCA), latent semantic analysis (LSA), and random projection (RP) feature selection methods are presented in this paper. Furthermore, the hybrid sentiment classification was validated using Twitter datasets to represent different domains, and the evaluation with different classification algorithms also demonstrated that the new hybrid approach produced meaningful results. The implementations showed that the new hybrid sentiment classification was able to improve the accuracy performance from the existing baseline sentiment classification methods by 76.55, 71.62 and 74.24%, respectively.

Keywords Aspect-based sentiment classification · Feature selection · Principal component analysis · Support vector machine · Hybrid approach

✉ Ali Selamat
  aselamat@utm.my

  Nurulhuda Zainuddin
  nurulhuda066@melaka.uitm.edu.my

  Roliana Ibrahim
  roliana@utm.my

[1] Faculty of Computing, Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310 Johor, Malaysia

[2] MAGIC-X (Media and Game Innovation Centre of Excellence), Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310 Johor, Malaysia

[3] Centre for Basic and Applied Research, Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, Hradec Kralove, 500 03, Czech Republic

1 Introduction

The evolution of the social media web and the chance to access the important opinions of different people on various business, political, health and social issues have motivated the development of sentiment analysis as a dynamic and important research field [11]. According to [10], a document-level sentiment analysis can be defined as the simplest form of sentiment analysis, where it is assumed that the document contains an opinion on one main object expressed by the author of the document. A document may even contain multiple opinions about the same entity. On the other hand, a sentence-level sentiment analysis is a detailed form of sentiment analysis aimed at obtaining a more detailed view of the different opinions expressed in the document about the entities. However, it is necessary to determine if the sentences are subjective or objective before the polarity of the sentences can be analysed. Only the subjective sentences will then be further analysed to determine whether they are positive, negative or neutral [10].

A sentiment analysis (SA) of texts at the document level or the sentence level is always not enough for applications because it does not specifically have an opinion target; in other words, the sentiment does not have a target entity. An entity normally has many aspects, and people have different opinions about each of these aspects. For instance, when people talk about a product (entity), they may consider many aspects of the product such as its price, colour, weight, etc. Thus, if it is assumed that each document represents a single entity, a positive opinion in a document about the entity does not show that people have positive opinions about all aspects of the entity. In the same way, a negative opinion in the document does not mean that people are negative about all aspects of the entity [21]. Therefore, a complete analysis must be conducted in order to determine the possible aspects and to identify whether the sentiment about an aspect is positive or negative. A full aspect-based sentiment analysis (ABSA) model is needed to extract the sentiments in detail, especially in reviews of such products as cameras and smart phones, as well as specific brands such as Apple, Samsung, Google, etc. Other relevant sources are discussion forums where people give reviews and share their experiences in using the product.

Basically, an aspect-based sentiment analysis (ABSA) performs a finer-grained analysis, which is also defined as a research problem that focuses on identifying the sentiment expressions of aspects of the target within a given document [10]. Moreover, an ABSA aims at identifying the aspects of entities in the document, and for each identified aspect, the sentiment polarity is estimated based on a specific approach [7]. Most significantly, sentiment analyses at the document level and the sentence level do not determine exactly what people liked or did not like.

An aspect-based sentiment analysis (ABSA) consists of two main tasks: aspect-based feature extraction, and aspect-based sentiment classification [22]. Aspect-based feature extraction is the task where the main aspects of entities in a specific domain are identified. Most of research works on aspect-based feature extraction concentrated on nouns, noun phrases/groups [17–19, 23]. Other approaches used phrase dependency parser to take into consideration the relationship between aspects and opinions, as well as machine learning techniques, such as conditional random field (CRF), to find explicit aspects.

Other researchers proposed an unsupervised model consisting of methods for learning multi-word aspects of product reviews[6]. It considered the influence of an opinion word on detecting the aspect by employing a set of heuristic rules. Furthermore, the researchers proposed a new measurement based on mutual information and aspect frequency to score aspects with a recently developed bootstrapping iterative algorithm. Not all aspects detected are useful aspects and there are also some incorrect ones. Therefore the model uses aspect pruning to remove these incorrect aspects. Another study [20] proposed a pattern-based bootstrapping algorithm to extract candidate product features, and feature clustering to group the features into aspects. However, it did not handle more types of features such as adjectives and verbs, and did not consider implicit features. Besides, in 2013, authors [25] proposed two novel APSM and ME-APSM models to extract aspects and aspect-specific polarity-aware sentiments from online reviews. However, the results showed that the model still needed improvements in terms of the aspect-level sentiment classification.

The aim of our research is to propose a new hybrid sentiment classification approach using Twitter attributes as features to improve the Twitter aspect-based sentiment analysis performance. The hybrid model was validated using Twitter datasets, namely, the HCTS dataset [36], STS dataset [12], and Sanders Twitter Corpus (STC) dataset [29]. Our hybrid sentiment classification model incorporates rule-based with feature selection methods for the Twitter sentiment classification. The results were compared with a baseline classification method, i.e. the support vector machine. The main contributions of our research are outlined as follows:

(i) We examined whether the association rule mining (ARM) augmented with heuristic combination POS patterns is beneficial for detecting single and multi-word aspects for a Twitter aspect-based sentiment analysis, which has been studied extensively.

(ii) We proposed a new hybrid sentiment classification for a Twitter aspect-based sentiment analysis, which incorporates rule-based with feature selection methods, including principal component analysis (PCA), latent semantic analysis (LSA), and random projection (RP) during the experiments.

(iii) The proposed dataset, named Hate Crime Twitter Sentiment (HCTS), was released for evaluation by other researchers.

The remaining part of the paper proceeds as follows: Section 2 describes related works on sentiment analysis, Section 3 explains the proposed hybrid sentiment classification framework, whilst Section 4 describes the experimental results and discussion. Then, the final section presents the conclusions and future works.

2 Related works on sentiment analysis

In recent years, a large number of techniques and enhancements have been proposed for the problem of sentiment analysis (SA) in different fields and for different tasks. Three types of techniques are used to classify opinions in SA, namely, lexicon-based approaches, machine learning approaches and hybrid approaches.

Lexicon-based approaches do not require any training dataset and use sentiment lexicons such as Word Net [24] and SentiWordNet [5], for classification purposes. These approaches give sentiment scores ranging from −1 to 1, but they do not classify the context-dependent opinion words appropriately. Moreover, there are also hybrid approaches that combine machine learning and lexical approaches [4].

In contrast, machine learning approaches can be grouped into two main categories, which are supervised and unsupervised techniques. Support Vector Machine (SVM) and Naive Bayesian (NB) classifications are examples of supervised sentiment analysis techniques which have achieved higher success in text classifications [26, 27, 33]. A primary concern of supervised approaches is that they depend on large training datasets, which can be time consuming to collect for each domain. The supervised methods also depend on the selection and extraction of the appropriate set of features to be used for the detection of sentiments. For instance, unigrams, bigrams, and part-of-speech tags are used as feature extractors [26, 31]. Feature selection enhances the sentiment classification method by combining syntactic features with semantic information from sources like SentiWordNet [1]. Feature selection methods, such as Principal Component Analysis (PCA) and Random Projection (RP), are aimed at eliminating irrelevant and redundant features to yield an improved classification accuracy for machine learning techniques and also to reduce data dimensionality [28]. Various feature selection approaches, such as information gain and the chi-square test, are employed to gain higher accuracy in sentiment analysis [35], but many of these studies mainly focused on document-level sentiment analysis.

Authors [3] presented an efficient method of feature selection and ensemble learning for an aspect-based sentiment analysis. The algorithm is based on a single-objective PSO and basic learning algorithms, namely CRF, SVM and ME. Using the SVM+PCA method, an accuracy of 74.51% was obtained on a laptop, on restaurant review datasets. The aim of this works was to investigate the use of a hybrid sentiment classification to enhance the performance of a Twitter aspect-based sentiment classification.

3 Proposed hybrid sentiment classification framework

Figure 1 shows Twitter aspect-based hybrid sentiment classification framework. By using this framework, it can obtain highly effective results for hybrid sentiment classification. The framework performs sentiment classification in four main phases: data collection using twitter datasets, twitter preprocessing incorporates filtering to filter unique twitter attributes, aspect-based feature extraction to identify single and multi-word explicit and implicit aspects by using an association rule mining (ARM) with heuristic combination POS patterns and Stanford dependency parser (SDP) methods. Aspect-based hybrid sentiment classification consists of rule-based approach for sentiment word detection and principal component analysis (PCA) for sentiment word feature selection and lastly sentiment classification using support vector machine (SVM).

3.1 Data collection

Table 1 shows the Hate Crime Twitter Sentiment (HCTS) dataset [36]. This dataset consists of different categories of hate crimes including racial, religion, sexual, feminist, disability and nationality. Table 2 shows the Stanford Twitter Sentiment (STS) dataset [12]. We considers only 353 tweets of negative and positive based on different categories. Table 3 shows the Sanders Twitter Corpus (STC) dataset

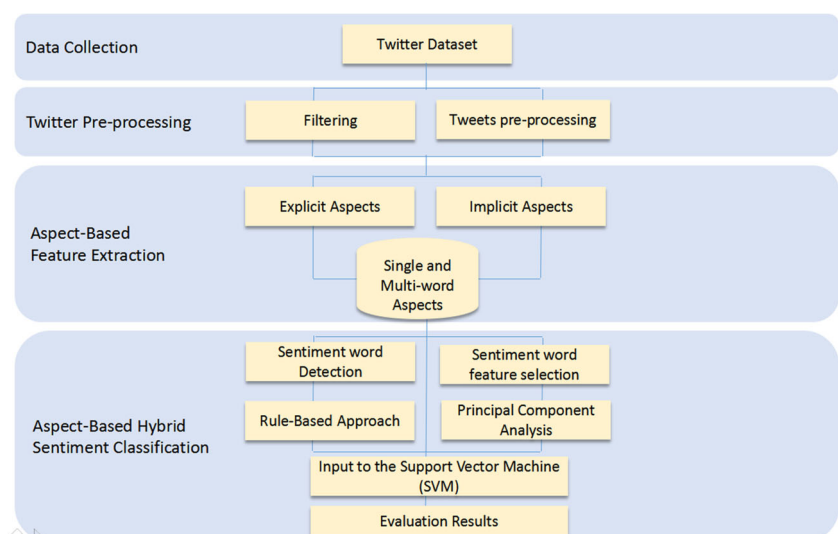**Fig. 1** Twitter aspect-based hybrid sentiment classification framework

Table 1 Hate Crime Twitter Sentiment (HCTS) Dataset

Dataset	Category C_h	Sample Query String	No of tweets analysed n_d
	race	anti-black,anti-white	255
HCTS Dataset	religion	anti-Muslim, anti-Jews	253
	sexual	lesbian, sorient	100
	feminist	anti-feminist	114
	disability	cripple, disability	197
	nationality	xenophobia	159
	Total		1078

[29] which can be downloaded from http://www.sananalytics.com/lab/twittersentiment/. This dataset consists of four different categories (Apple, Google, Microsoft, and Twitter), and has 5513 manually-classified positive, negative, neutral and irrelevant tweets. In this experiments, we used only 519 positive and 572 negative tweets whilst the neutral and irrelevant tweets which were not necessary for the classification, were ignored [37]. These datasets were chosen because it were used by other researchers as an evaluation dataset for the twitter sentiment analysis[29]. Using this datasets proves that our proposed approach is domain-independent for detecting explicit and implicit aspects in twitter sentiment analysis in order to attain highly effective results.

3.2 Text pre-processing

The tweets had to go through the pre-processing steps prior to the classification because the language of Twitter has some unique attributes that may not be relevant to the classification process, such as usernames, links and hashtags. Besides, the filtering process was carried out by removing new lines, opposite emoticons, repeated letters, laughter and punctuation marks. Finally, tokenization, the removal of stop words, lowercase conversion, and the stemming process were carried out to complete the twitter pre-processing.

Table 2 Stanford Twitter Sentiment (STS) Dataset

Dataset	Category C_h	Sample Query String	No of tweets analysed n_d
	company	aig,goodby silverstein	116
STS Dataset	event	world cup, indian election	8
	location	north korea, san francisco	17
	misc.	insects, exam	66
	movie	night at the museum	19
	person	obama, ,malcolm gladwell, cheney	64
	product	nike	63
	Total		353

Table 3 Sanders Twitter Corpus (STC) Dataset

Dataset	Category C_h	Sample Query String	No of tweets analysed n_d
	Apple	apple	480
STC Dataset	Google	google	259
	Microsoft	microsoft	223
	Twitter	twitter	129
	Total		1091

3.3 Aspect-based feature extraction

The task was performed to find the explicit single and multi-word aspects by using the association rule mining (ARM) with heuristic combination POS patterns. In addition, the Stanford Dependency Parser (SDP) method, which took into account the relationship between the opinion and aspects, was employed for the extraction of implicit aspects. These elements are discussed in the following sub sections.

3.3.1 Single and multi-word explicit aspect extraction using Association Rule Mining (ARM)

Association rule mining is used to find the important aspects of a given target. Association rules are created by analysing the data for frequent 'if/then' patterns. Then, the supporting criteria and confidence are used to identify the most important relationships. In other words, the confidence indicates the number of times the 'if/then' statements have been found to be true, whilst the supporting criteria are an indication of how frequently the items appear in the database.

The association rule mining is stated as follows:

Let $I = i_1, ..., i_n$ be a set of items, and D be a set of transactions (the dataset). Each transaction consists of a subset of items in I. An *association rule* is an implication of the form $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule $X \rightarrow Y$ holds in D with confidence c if $c\%$ of the transactions in D that support X also support Y. The rule has support s in D if $s\%$ of the transactions in D contain $X \cap Y$. The problem with association rule mining is that it has to generate all the association rules in D that have support and confidence that are greater than a user-specified minimum support and minimum confidence.

We used association rule mining (ARM) to find the important single and multi-word aspects for the given target entities [13]. In our case, an aspect was defined as being important if it appeared in more than 1% (minimum support) of the sentences. This experiment applied the association rule mining which is based on the Apriori algorithm. The Apriori algorithm is used to find the *frequent (important) aspects* from a set of transactions that satisfy a user-specified minimum support.

Table 4 Alphabetical list of part-of-speech (POS) tags used in the Penn Treebank Project for ARM Method

Tag	Descriptions
DT	Determiner
JJ	Adjective
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present

The different values for the minimum support and minimum confidence were applied in the experiments. During the experiments, a suitable value for the minimum support was 0.1, and for the minimum confidence it was 0.5.

From the first experiments of single and multi-word explicit aspect extraction using association rule mining (ARM), a list of single aspects was obtained from the nouns and noun phrases. In this phase, the heuristic combination POS patterns were applied in order to identify the multi-word aspects from the tweets. Previous research had established that a few aspects that people talk about have more than one single word, especially in review sentences [6]. Table 4 shows an alphabetical list of the part-of-speech (POS) tags used in the Penn Treebank Project for the ARM method, whilst Table 5 shows the heuristic combination

Table 5 Heuristics combination in POS patterns for multi-word aspects generation

Description of features	Patterns
Nouns	Unigram, Bigram, Trigram of NN, NNS, NNP, NNPS
Nouns and Adjectives	Bigram to Trigram of [NN JJ], [NNS JJ], [NNP JJ],[NNPS JJ]
Determiners and Adjectives	Bigram of [DT JJ]
Nouns and Verbs	Bigram to Trigram of [NN VB],[NN VBD],[NN VBG],[NN VBP],[NN VBZ]
Nouns and Adverbs	Bigram to Trigram of [NN RB],[NN RBR],[NN RBS]

part-of-speech (POS) patterns for the aspects generation. In this experiment, for instance, the phrases "feminism", "racism", and "blackness" were obtained, which were extracted from the pattern "NN NNS NNP NNPS". Then, another phrase, "white racist" was extracted from the "NN JJ" pattern, "anti bigotry" was extracted from the "NN VBG" pattern, "anti-muslim" was extracted from the "NN VBZ" pattern, and "anti racist" was extracted from the "NN RBR" POS pattern.

Below is the sample of multi-word aspects generation obtained from heuristic combination in POS patterns for HCTS, STS and STC datasets:

- (NN—JJ), e.g. aspects from the HCTS dataset: anti white, anti blackness, anti feminist, anti america, anti muslim; aspects from the STS dataset: night museum, malcolm gladwell, amp t, time warner; aspects from the STC dataset: ice cream sandwich.
- (NN—VB) e.g. anti white supremacy.
- (DT—JJ) e.g. anti muslim prejudice.

3.3.2 Implicit aspects extraction using the Stanford Dependency Parser (SDP) method

Few studies have investigated the implicit aspects extraction for an aspect-based sentiment analysis. In this work, we attempted to show that the relationship between aspects and opinions can help to determine implicit aspects by capturing the grammatical relations by using the dependency parsers. Different types of dependencies were used to find the relationships that were beneficial to discover implicit aspects. Table 6 shows a sample description of the types of dependencies that were used during the implicit aspects extraction [8, 9].

Table 6 Sample description of typed dependency

Typed Dependency	Grammatical Relation Description
det	determiner - specific reference to a noun phrase
amod	adjectival modifier - descriptive phrase related to a noun phrase
aux	auxiliary - a form of 'be','do' or 'have' - action phrase
dobj	direct object - an accusatory object of the verb
advmod	adverbial modifier - descriptive phrase related to a verb
neg	negation modifier - the relation between a negation word and the word it modifies.
nsubj	nominal subject - is a noun phrase which is the syntactic subject of a clause.
xcomp	open clausal complement of a verb or an adjective is a predictive or clausal complement without its own subject.

In this process, take, for instance, this tweet:

Example 1 "Govt removing Motability cars from disabled ppl who can only walk 50m and will then demonise those that have to give up work because of it."

The SDP method yielded the following results:

root(ROOT-0, Govt-1) acl(Govt-1, removing-2) compound(cars-4, Motability-3) dobj(removing-2, cars-4) case(ppl-7, from-5) amod(ppl-7, disabled-6) nmod (removing-2, ppl-7) nsubj(walk-11, who-8) aux(walk-11, can-9) advmod(walk-11, only-10) acl:relcl(ppl-7, walk-11) dobj(walk-11, 50m-12) cc(walk-11, and-13) aux(demonise-16, will-14) advmod(demonise-16, then-15) conj(walk-11, demonise-16) dobj(demonise-16, those-17) nsubj(have-19, that-18) acl:relcl(those-17, have-19) mark(give-21, to-20) xcomp(have-19, give-21) compound:prt(give-21, up-22) dobj(give-21, work-23) case(it-26, because-24) mwe(because-24, of-25) nmod(work-23, it-26).

Example 2 "People that openly laugh at handicapped people in public are a disgrace. So many pathetic people on this planet"

The SDP method yielded the following results:

nsubj(disgrace-12, People-1) nsubj(laugh-4, that-2) advmod(laugh-4, openly-3) acl:relcl(People-1, laugh-4) case(people-7, at-5) amod(people-7, handicapped-6) nmod(laugh-4, people-7) case(public-9, in-8) nmod(people-7, public-9) cop(disgrace-12, are-10) det(disgrace-12, a-11) root(ROOT-0, disgrace-12) root(ROOT-0, So-1) amod(pathetic-3, many-2) amod(people-4, pathetic-3) nmod(So-1, people-4) case(people-4, on-5) det(planet-7, this-6) dep(people-4, planet-7).

The process used direct dependencies and transitive dependency (within a distance of one dependency relation). From Example 1, an implicit aspect was identified from the tweet, namely, 'disabled people' from the relation amod(ppl-7, disabled-6)'. Besides, another implicit aspect was identified from the relation amod(people-7, handicapped-6), which was taken from Example 2. 'Disabled people' and 'handicapped people' were implicit aspects that actually referred to the 'disability hate crime' category. Besides, when a negation modifier relation was found, it would change the tweet to the opposite sentiment.

3.4 Aspect-based hybrid sentiment classification

The hybrid sentiment classification contains a rule-based and a Feature Selection (FS) for identifying the sentiment words, as well as Support Vector Machine for sentiment classification.

3.4.1 Rule-based method for sentiment word detection

For the purpose of sentiment word detection, the rule-based method was employed. It contains two main rules, as follows:

```
  FOR each sentence in the tweets
Read the aspect in the sentence
IF the aspect match THEN
Get sentimentword
IF sentimentword distance less than or equal
to 4
Add the sentimentword into Result
ELSE
Remove sentimentword
END IF
ELSE Display no sentiment word
END IF
END FOR

  FOR each sentimentword in the Result
Compute the sentimentwordvalue
Display sentimentwordvalue
END FOR
```

For example, in this tweet: *"If I can help myself I don't need you, I'm still blessed to have my limbs so I will use them. I hate when people make me feel disabled!"*
The SDP method successfully extracted an aspect from the relation "xcomp - open clausal complement of a verb or an adjective", which was "feel disabled", from tweets. However, not all the aspects identified in the first process would be considered as significant aspects. The role of the rule-based method is to identify the significant aspects. It is based on the location of the aspects and sentiment word in the tweets, and also the polarity value for every tagged sentiment word in the tweets.

3.4.2 Sentiment word feature selection

The Principal Component Analysis (PCA) algorithm can be used in feature selection method. It has previously been observed that the PCA showed promising results in the feature selection process [30, 32].

The PCA-based dimensionality reduction is based on the following steps: 1) Convert training and test datasets into numerical form; 2) Find covariance matrix of datasets; 3) Calculate Eigen values and Eileen vector of the covariance matrix; 4) Sort Eigen vectors w.r.t non-increasing Eigen values; 5) Keep the top k vectors; and 6) Train, test and evaluate the reduced datasets.

3.4.3 Baseline classification methods

It is now well-established from a variety of studies [2, 16, 26] that a Support Vector Machine (SVM) works well for

Table 7 Some experimental results of aspects detection for HCTS dataset

Category	List of single aspects based on ARM	List of multi-aspects based on ARM + Heuristic Combination POS Patterns
Racial	racism, racist, blackness, supremacy, blacks, hate, whites, race, violence, cop	black people, black person, black liberation, south africa, white privilege, white supremacy, anti-white racism, anti-white racist
Religion	islamophobia, israel, propaganda, jews, islam, muslim, jewish, bigotry, sentiment, racism, hatred, arab, zionist, racist, semitism, semitic, propaganda, violence, israeli	anti muslim, anti jewish, anti islam
	jewish	
Sexual/Feminist	feminist, anti, equality; anti, boys; anti, rights; anti, post; anti, men; anti, gender; anti, class; feminism; women; anti, girls, lesbian	Anti-feminist, gay people, gay marriage
Disability	cripple, disable	cripple girl, deaf people, disabled people, disabled child
Nationality	anti, chinese, america, russia, japan, sentiment,eurovis, russia, world, hate japanese, arab, xenophobia, racism, russia, eurovis, anti, world	south africans, anti russia, dominican republic, haitian descent anti-chinese sentiments

various classifications and text categorisations. The SVM has many advantages in terms of handling large features and robustness when there is a sparse set of examples, especially when only a small number of tweets are used in the experiments.

4 Experimental results and discussion

Three datasets were used for the experiment, including the STS dataset, which consists of companies, events, locations, miscellaneous, movies, persons and products; the HCTS dataset, which has six types of hate crime categories, namely, racial, religion, sexual, feminist, disability and nationality; and lastly, the STC dataset, which consists of four categories, namely Google, Microsoft, Twitter and Apple.

From the process, a list of single and multi-word aspects was obtained from the association rule method with heuristic combination POS patterns, as shown in Table 7, for the HCTS dataset. Each tweet was annotated with a list of aspects that were relevant to the dataset. Table 7 shows that the process was able to successfully extract aspects from the HCTS dataset, including "anti", "muslim", "feminist", "jewish", "black", and "white", which were relevant to the targets of this research. Furthermore, the results also shows the multi-word aspects that were extracted from the heuristic combination part-of-speech (POS) patterns.

Tables 8 and 9 present a summary of the aspects for the HCTS and STS datasets, respectively, based on the dependency parser grammatical relation.

4.1 Evaluation measures

Accuracy measures were used to evaluate the entire classification performance with binary classes (positive and negative). For the positive and negative sentiments on the entities, the standard evaluation measures of precision and recall were employed.

Four effective measures were used in this study based on the confusion matrix output, namely, True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

- Precision(P) = TP/(TP+FP)
- Recall(R) = TP/(TP+FN)
- Accuracy(A) = (TP+TN)/(TP + TN + FP + FN)
- F-Measure(Micro-averaging) = 2.(P.R)/(P+R)

4.2 Aspect based sentiment feature extraction analysis results

As can be seen from Table 10, the STS dataset gained higher positive polarity scores than negative polarity scores. The experimental results showed that the total percentage of positive labels was 21.6% from companies, 87.5%

Table 8 Implicit Aspects from Hate Crime Twitter Sentiment (HCTS) Dataset

Relation	Dependency Parsing
nsubj, amod, advmod, xcomp	anti-racist, anti-stupid, pro-happy,nigga nigger, anti-black, neo-nazism, anti-oppression, anti conservative, Israeli hate, sexism, anti indigeneity, anti-semitic, anti-indigenous, islamophobia, gender equality, reverse racism, anti equality, anti-Gay, anti-christ

Table 9 Implicit aspects from Stanford Twitter Sentiment (STS) dataset

Relation	Dependency Parsing
nsubj, amod, advmod, xcomp	AIGquality, NIKEflywire, NIKEcommercial, ,NIKEbasketball, NIKEtraining-clubapps, GOOGLEslide screen, GOOGLEplace, GOOGLEtranslator, GOOGLEvideo, MCDONALDeggs, MCDONALDmeal, TIMEWARNERcable, TIMEWARNERcustomerservice, TIMEWARNERdvr, LEBRONpersonality, LEBRONcharacter, OBAMAspeech, OBAMAfunnyside, OBAMAconservative

from events, 17.6% from locations, 28.8% from miscellaneous, 78.9% from movies, 65.6% from persons, and 54% from product categories. Furthermore, the total percentage of negative labels obtained was 61.2% from companies, 0% from events, 47.1, 42.4, 5.3, 18.8, and 20.6% from other categories, respectively. This method also successfully classified tweets as neutral, where it obtained 17.2% from companies, 12.5% from events, 35.3% from locations, 28.8% from miscellaneous, 15.8% from movies, 15.6% from persons, and 25.4% from product categories.

The results obtained from the feature extraction analysis of the HCTS dataset are presented in Table 11. The results were classified as positive, negative and neutral tweets. Therefore, the disability, racial, feminist and religious categories gained higher negative polarity scores compared to positive polarity scores. In contrast, the sexual category gained higher positive polarity scores of 27% compared to negative polarity scores of 25%. The proposed method labelled tweets as negative for 27.9% from the disability category, 57.9% from the feminist category, 55.3% from the racial category, and 49.5% from the religious category. Nevertheless, 21.3%, 16.7%, 16.5% and 13.4% of the tweets, respectively from the HCTS dataset were labelled as positive.

4.3 Twitter aspect-based hybrid sentiment classification analysis results

Tables 12, 13 and 14 present the performances of the aspect-based classifier with different feature selection methods, respectively. Here, the methods considered only the subjective tweets. Then, the performance was measured using the accuracy, precision, recall and F-measure. The evaluation was carried out with a combination of different methods

to be incorporated with the SVM classifier. Three feature selection methods were used, namely, principal component analysis (PCA), latent semantic analysis (LSA), and random projection (RP).

Table 12 presents the experimental results from the STS dataset. It shows that the highest accuracy achieved was 76.5517 with the $ABSA + Sentiwordnet + PCA$ method. Besides, the precision was 0.779, recall 0.766, and F-measure 0.76 with POS Tags features. In contrast, by using the ABSA+Sentiwordnet alone, the accuracy achieved was only 53.4483, which was the same when the LSA method was applied. Likewise, by using the random projection (RP) method, the highest accuracy achieved was only 63.7931%, with a combination of POS Tags and bigram features. The results have been presented in Fig. 2 that shows STS classification results between different features and feature selection method.

In contrast, the experimental results from the HCTS dataset presented in Table 13 showed that the proposed approach using the $ABSA + Sentiwordnet + PCA$ method with a combination of POS Tags and unigram features gave the highest accuracy of 71.6243%, 0.708 for precision, 0.716 for recall, and 0.647 for the F-measure, accordingly. On the other hand, by using the $ABSA + Sentiwordnet$ with POS Tags features alone, it gave an accuracy of only 69.0802% compared to the $ABSA + Sentiwordnet + PCA$ with POS Tag features, which obtained an accuracy of 71.2329%. Lastly, by using the $ABSA + Sentiwordnet + RP$ with a combination of POS Tags + unigram features, the accuracy achieved was 70.8415%. Furthermore, the PCA method was compared with the latent semantic analysis (LSA) feature selection method, where the results indicated that the accuracy did not change with the use of the same

Table 10 STS polarity scores and total percentage of positive, negative and neutral labels according to category

Category	Positive Polarity Scores	Negative Polarity Scores	Positive %	Negative %	Neutral %
Company	12.75	33.5	21.6%	61.2%	17.2%
Event	4.5	0	87.5%	0%	12.5%
Location	1.75	2.75	17.6%	47.1%	35.3%
Misc	10.25	14.125	28.8%	42.4%	28.8%
Movie	9.75	0.875	78.9%	5.3%	15.8%
Person	23.75	6.5	65.6%	18.8%	15.6%
Product	21	5.5	54%	20.6%	25.4%

Table 11 HCTS polarity scores and total percentage of positive, negative and neutral labels according to category

Category	Positive Polarity Scores	Negative Polarity Scores	Positive %	Negative %	Neutral %
Racial	6.75	43.375	16.5%	55.3%	28.2%
Religion	5.625	18.625	13.4%	49.5%	37.1%
Sexual	8.75	10.75	27%	25%	48%
Feminist	7.5	21.625	16.7%	57.9%	25.4%
Disability	13.125	21.917	21.3%	27.9%	50.8%
Nationality	5.375	34.042	9.43%	41.5%	49.06%

Table 12 Comparison of STS aspect-based sentiment classification results with different feature selection methods for 10-fold cross validation

Method	Features	Accuracy(%)	Precision	Recall	F-Measure
ABSA+Sentiwordnet+SVM	POS Tags	53.4483	0.286	0.534	0.372
ABSA+Sentiwordnet+PCA+SVM	POS Tags	**76.5517**	**0.779**	**0.766**	**0.76**
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Unigram	70.6897	0.732	0.707	0.693
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Bigram	63.1034	0.633	0.631	0.622
ABSA+Sentiwordnet+LSA+SVM	POS Tags	53.4483	0.286	0.534	0.372
ABSA+Sentiwordnet+LSA+SVM	POS Tags+Unigram	53.4483	0.286	0.534	0.372
ABSA+Sentiwordnet+LSA+SVM	POS Tags+Bigram	53.4483	0.286	0.534	0.372
ABSA+Sentiwordnet+RP+SVM	POS Tags	54.1379	0.535	0.541	0.529
ABSA+Sentiwordnet+RP+SVM	POS Tags+Unigram	58.279	0.58	0.583	0.578
ABSA+Sentiwordnet+RP+SVM	POS Tags+Bigram	63.7931	0.658	0.638	0.613

The bold emphasis show the highest accuracy achieved by the methods proposed

Table 13 Comparison of HCTS aspect-based sentiment classification results with different feature selection methods for 10-fold cross validation

| Method	Features	Accuracy(%)	Precision	Recall	F-Measure
ABSA+Sentiwordnet+SVM	POS Tags	69.0802	0.477	0.691	0.564
ABSA+Sentiwordnet+PCA+SVM	POS Tags	71.2329	0.694	0.712	0.646
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Unigram	**71.6243**	**0.708**	**0.716**	**0.647**
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Bigram	67.7104	0.608	0.677	0.601
ABSA+Sentiwordnet+LSA+SVM	POS Tags	69.0802	0.477	0.691	0.564
ABSA+Sentiwordnet+LSA+SVM	POS Tags+Unigram	69.0802	0.477	0.691	0.564
ABSA+Sentiwordnet+LSA+SVM	POS Tags+Bigram	69.0802	0.477	0.691	0.564
ABSA+Sentiwordnet+RP+SVM	POS Tags	70.2544	0.681	0.703	0.617
ABSA+Sentiwordnet+RP+SVM	POS Tags+Unigram	70.8415	0.724	0.708	0.615
ABSA+Sentiwordnet+RP+SVM	POS Tags+Bigram	69.4716	0.654	0.695	0.601

The bold emphasis show the highest accuracy achieved by the methods proposed

Table 14 Comparison of STC aspect-based sentiment classification results with different feature selection methods for 10-fold cross validation

Method	Features	Accuracy(%)	Precision	Recall	F-Measure
ABSA+Sentiwordnet+SVM	POS Tags	52.429	0.275	0.524	0.361
ABSA+Sentiwordnet+PCA+SVM	POS Tags	73.3272	0.744	0.733	0.728
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Unigram	**74.2438**	**0.751**	**0.742**	**0.738**
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Bigram	59.5784	0.707	0.596	0.517
ABSA+Sentiwordnet+LSA+SVM	POS Tags	52.429	0.275	0.524	0.361
ABSA+Sentiwordnet+LSA+SVM	POS Tags+Unigram	52.429	0.275	0.524	0.361
ABSA+Sentiwordnet+LSA+SVM	POS Tags+Bigram	52.429	0.275	0.524	0.361
ABSA+Sentiwordnet+RP+SVM	POS Tags	67.0944	0.681	0.671	0.662
ABSA+Sentiwordnet+RP+SVM	POS Tags+Unigram	63.703	0.64	0.637	0.631
ABSA+Sentiwordnet+RP+SVM	POS Tags+Bigram	59.5784	0.71	0.596	0.516

The bold emphasis show the highest accuracy achieved by the methods proposed

Fig. 2 STS classification results between different features and feature selection method

STS COMPARISON METHOD FOR ASPECT-BASED SENTIMENT CLASSIFICATION RESULTS WITH 10-FOLD CROSS VALIDATION
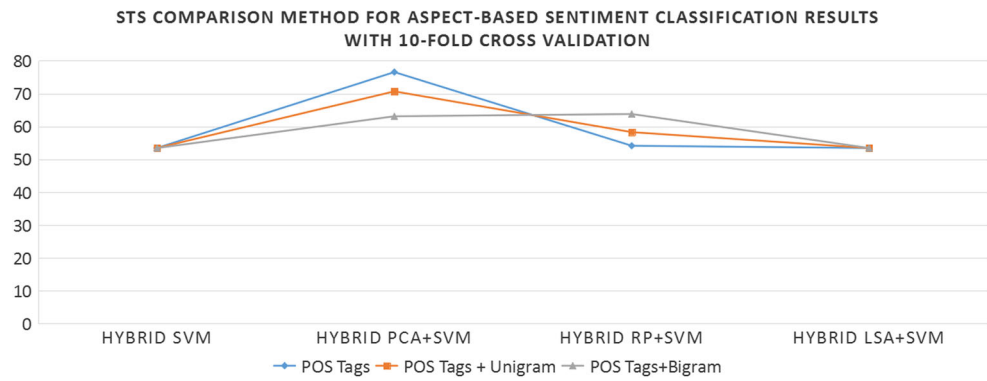
features. Figure 3 shows detail HCTS classification results between different features and feature selection method.

Table 14 shows the experimental results from the STC dataset. It also revealed that the highest accuracy was achieved by using the $ABSA + Sentiwordnet + PCA$ method, where the accuracy was 74.2438, with a combination of POS Tags and unigram features. Besides, the precision shown was 0.751, recall was 0.742 and F-measure was 0.738. In contrast, by using the $ABSA + Sentiwordnet$ method alone, the accuracy achieved was only 52.429, which was the same as when the LSA method was applied. It was apparent from this table that the feature selection with random projection (RP) method could only produce an accuracy of 67.0944% with POS Tags features. It can be seen in Fig. 4 the detail results of STC dataset with different features and feature selection method. From the graphs in Figs. 2, 3 and 4, it can be seen that the part-of-speech (POS) tags were the best features for the feature selection method.

4.4 Discussion I

The results of this study were then compared with the findings from different feature selection and classification algorithms, as shown in Table 15, 16 and Table 17. The F-measure of 0.76 from the STS dataset was obtained by using the hybrid $ABSA + Sentiwordnet + PCA + SVM$ approach, and was followed by 0.729 with the random forest (RF) classifier. A comparison of the results from the HCTS dataset revealed that the highest F-measure value of 0.647 was also obtained from the hybrid $ABSA + Sentiwordnet + PCA + SVM$ approach.

The results of this experiment were then compared with the findings from the STC dataset, where the same approach obtained the highest F-measure value of 0.738. Surprisingly, as can be seen in Table 17, the hybrid $ABSA + Sentiwordnet + PCA + RF$ method successfully achieved an F-measure value of 0.74, which was closer to the result with the SVM. However, a longer time was taken to train the random forest (RF) classifier compared to the other classification algorithms.

The classification accuracies of the STS, HCTS and STC datasets were also better with the $ABSA + Sentiwordnet + PCA + SVM$ approach compared to the other classifiers. An accuracy of 76.5517% was achieved with the STS, 71.6243% with the HCTS, and 74.2438% with the STC datasets. In summary, these results showed that the Support

Fig. 3 HCTS classification results between different features and feature selection method

HCTS COMPARISON METHOD FOR ASPECT-BASED SENTIMENT CLASSIFICATION RESULTS WITH 10-FOLD CROSS VALIDATION
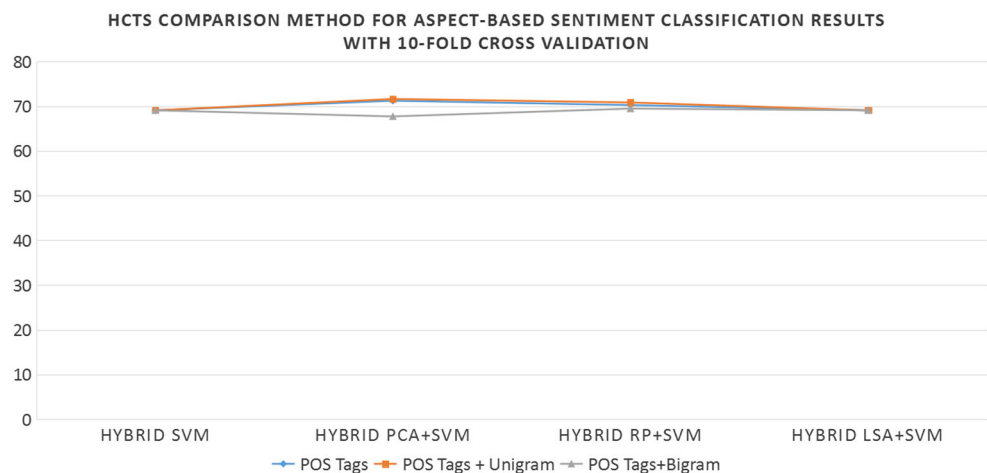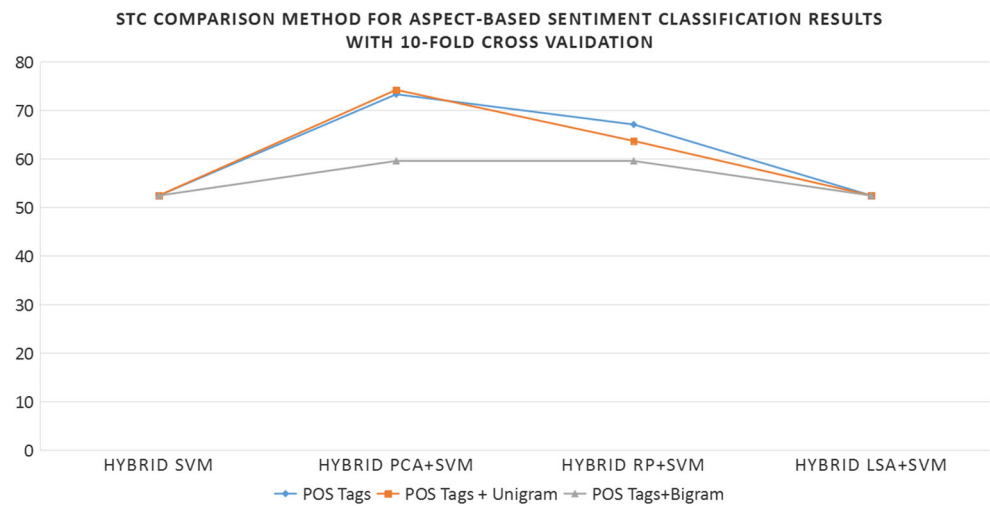
Fig. 4 STC classification results between different features and feature selection method

Vector Machine (SVM) worked well even though different features were used during the classification, and that the part-of-speech (POS) tags feature was the best feature for representing the tweets during the classification. Overall, these results also indicated that the classification accuracies were improved with the proposed hybrid approach.

4.5 Discussion II

The performance of the extreme learning machine (ELM) has been tested in the STS, HCTS and STC datasets in the binary classification, which has two training samples in each class. An analysis by Huang [14] identified the ELM as being efficient, accurate and easy to implement in various applications. The aim of this experiment was to verify whether the ELM can handle a few training datasets, especially Twitter datasets. The experiments were conducted on a laptop with a sigmoidal hidden layer activation function. The number of hidden nodes, L varied from 20 to 500, with an interval of 10 [34]. Fifty trials were conducted for each problem in the experiment [15]. Table 18 describes the

performance comparison of ELM for the binary class Twitter dataset. The table includes the average testing and training accuracy, and also the corresponding standard deviation (Std Dev).

Table 18 highlights in particular, the performance comparison between the testing and training accuracies with various hidden nodes from [20, 500]. The best performance of the hybrid method was produced by the $ABSA + Sentiwordnet + ELM + Unigram$ for training the STS dataset, where the training accuracy was 0.95249. Besides, for the HCTS dataset, the best performance of training accuracy was produced by the $ABSA + Sentiwordnet + ELM + Unigram$, where the value was 0.92623. In addition, for the STC dataset, the best performance of the hybrid method was also produced by the $ABSA + Sentiwordnet + ELM + Unigram$, where the value was 0.84126.

The testing results of the datasets were compared with the training results. The best performances of the hybrid method were produced in the HCTS and STC datasets. It could be seen that the number of samples in the testing and training datasets played a significant role in producing

Table 15 Evaluation of STS twitter aspect-based sentiment classification results with different classification algorithms

Method	Features	Accuracy(%)	Precision	Recall	F-Measure
ABSA+Sentiwordnet+PCA+SVM	POS Tags	**76.5517**	**0.779**	**0.766**	**0.76**
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Unigram	70.6897	0.732	0.707	0.693
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Bigram	63.1034	0.633	0.631	0.622
ABSA+Sentiwordnet+PCA+NB	POS Tags	66.8966	0.676	0.669	0.669
ABSA+Sentiwordnet+PCA+NB	POS Tags+Unigram	62.4138	0.67	0.624	0.61
ABSA+Sentiwordnet+PCA+NB	POS Tags+Bigram	67.2414	0.732	0.672	0.659
ABSA+Sentiwordnet+PCA+RF	POS Tags	73.1034	0.735	0.731	0.727
ABSA+Sentiwordnet+PCA+RF	POS Tags+Unigram	73.4483	0.743	0.734	0.729
ABSA+Sentiwordnet+PCA+RF	POS Tags+Bigram	66.5517	0.667	0.666	0.66

The bold emphasis show the highest accuracy achieved by the methods proposed

Table 16 Evaluation of HCTS twitter aspect-based sentiment classification results with different classification algorithms

Method	Features	Accuracy(%)	Precision	Recall	F-Measure
ABSA+Sentiwordnet+PCA+SVM	POS Tags	71.2329	0.694	0.712	0.646
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Unigram	**71.6243**	**0.708**	**0.716**	**0.647**
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Bigram	67.7104	0.608	0.677	0.601
ABSA+Sentiwordnet+PCA+NB	POS Tags	63.0137	0.612	0.63	0.62
ABSA+Sentiwordnet+PCA+NB	POS Tags+Unigram	66.1448	0.64	0.661	0.647
ABSA+Sentiwordnet+PCA+NB	POS Tags+Bigram	62.4266	0.597	0.624	0.607
ABSA+Sentiwordnet+PCA+RF	POS Tags	69.4716	0.654	0.695	0.601
ABSA+Sentiwordnet+PCA+RF	POS Tags+Unigram	71.2329	0.72	0.712	0.628
ABSA+Sentiwordnet+PCA+RF	POS Tags+Bigram	67.5147	0.617	0.675	0.614

The bold emphasis show the highest accuracy achieved by the methods proposed

Table 17 Evaluation of STC twitter aspect-based sentiment classification results with different classification algorithms

Method	Features	Accuracy(%)	Precision	Recall	F-Measure
ABSA+Sentiwordnet+PCA+SVM	POS Tags	73.3272	0.744	0.733	0.728
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Unigram	**74.2438**	**0.751**	**0.742**	**0.738**
ABSA+Sentiwordnet+PCA+SVM	POS Tags+Bigram	59.5784	0.707	0.596	0.517
ABSA+Sentiwordnet+PCA+NB	POS Tags	63.703	0.637	0.637	0.635
ABSA+Sentiwordnet+PCA+NB	POS Tags+Unigram	63.1531	0.631	0.632	0.629
ABSA+Sentiwordnet+PCA+NB	POS Tags+Bigram	58.5701	0.642	0.586	0.521
ABSA+Sentiwordnet+PCA+RF	POS Tags	71.5857	0.724	0.716	0.711
ABSA+Sentiwordnet+PCA+RF	POS Tags+Unigram	**74.2438**	**0.747**	**0.742**	**0.74**
ABSA+Sentiwordnet+PCA+RF	POS Tags+Bigram	59.5784	0.707	0.596	0.517

The bold emphasis show the highest accuracy achieved by the methods proposed

Table 18 Performance comparison in a benchmark dataset for Twitter aspect-based sentiment classification by using ELM

Dataset	Method and Features	Training Accuracy		Testing Accuracy	
		Mean	Std Dev	Mean	Std Dev
STS	ABSA+Sentiwordnet+ELM+POS Tags	0.95199	0.05871	**0.52017**	0.06743
	ABSA+Sentiwordnet+ELM+Unigram	**0.95249**	0.05975	0.51993	0.06582
	ABSA+Sentiwordnet+ELM+Bigram	0.80950	0.01786	0.51043	0.04241
HCTS	ABSA+Sentiwordnet+ELM+POS Tags	0.91955	0.08695	0.58397	0.05599
	ABSA+Sentiwordnet+ELM+Unigram	**0.92623**	0.09101	0.56276	0.05762
	ABSA+Sentiwordnet+ELM+Bigram	0.73547	0.00707	**0.71922**	0.02655
STC	ABSA+Sentiwordnet+ELM+POS Tags	0.82914	0.06063	0.61068	0.03610
	ABSA+Sentiwordnet+ELM+Unigram	**0.84126**	0.06184	**0.62823**	0.03331
	ABSA+Sentiwordnet+ELM+Bigram	0.61340	0.00000	0.55556	0.00457

The bold emphasis show the highest accuracy achieved by the methods proposed

Table 19 Paired sample T-test for classification performance measurements

Pair	*t*	*pValue*
Hybrid PCA+SVM—Hybrid PCA+NB	−4.781	0.004
Hybrid PCA+SVM—Hybrid PCA+RF	−0.6916	0.264
Hybrid PCA+SVM—Baseline SVM	−2.8386	0.023
Hybrid PCA+SVM—Baseline ELM	−2.0055	0.058

highly accurate results. A detailed analysis of the ELM for the aspect-based sentiment classification will be conducted for future works in this research.

4.6 Statistical significance tests

In this section, statistical tests were used to examine the significance of the differences in the means of the classification performances. Table 19 shows the results of the paired sample T-test for the classification performance measurements.

The statistical test results in Table 19 show that the differences in the hybrid methods were significant at an alpha level of 0.05. The difference between the hybrid PCA+SVM and hybrid PCA+NB was statistically significant at an alpha level of 0.05 with a p-value of 0.004. Moreover, the difference between the hybrid PCA+SVM and hybrid SVM was significant at an alpha level of 0.05 with a p-value of 0.023. However, the difference between the hybrid PCA+SVM and hybrid PCA+RF was statistically not significant with a p-value of 0.264, and it also showed that the difference between the hybrid PCA+SVM and hybrid ELM was statistically not significant, where the p-value was 0.058. However, it showed that the method was significant at an alpha level of 0.1.

5 Conclusions

In this paper, we proposed a new hybrid approach for a Twitter aspect-based sentiment analysis to perform finer-grained analysis. This research examined the association rule mining (ARM) that was augmented with a heuristics combination in part-of-speech (POS) patterns for detecting explicit single and multi-word aspects. The reason for this was that interrelations between a heuristic combination in POS patterns from words such as adjectives, adverbs, verbs and determiners with noun phrases are beneficial for the detection of relevant explicit aspects. Besides, the Stanford dependency parser (SDP) method, through the use of grammatical relations, is crucial in detecting implicit aspects. Our system also incorporates a rule-based with feature selection method for identifying the sentiment words. The evaluation with different classification algorithms also demonstrated that the new hybrid sentiment classification produced meaningful results with Twitter datasets, which represented different domains. The implementations showed that the new hybrid sentiment classification, that incorporated results from the aspect-based sentiment classifier method, was able to improve the performance of the existing baseline sentiment classification methods by 76.55, 71.62 and 74.24%, respectively. In a future work, we plan to conduct experiment with another social media data such as youtube and facebook by using the proposed hybrid sentiment classification approach in order to identify sentiment of people towards certain issues.

References

1. Abbasi A (2010) Intelligent feature selection for opinion classification. IEEE Intell Syst 25(4):75–79. https://www.scopus.com
2. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R (2011) Sentiment analysis of twitter data. In: Proceedings of the Workshop on Languages in Social Media, LSM '11. Association for Computational Linguistics, Stroudsburg, pp 30–38
3. Akhtar MS, Gupta D, Ekbal A, Bhattacharyya P (2017) Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. Knowl-Based Syst 125:116–135. https://doi.org/10.1016/j.knosys.2017.03.020. http://www.sciencedirect.com/science/article/pii/S095070511730148X
4. Appel O, Chiclana F, Carter J, Fujita H (2016) A hybrid approach to the sentiment analysis problem at the sentence level. Knowl-Based Syst 108:110–124. https://doi.org/10.1016/j.knosys.2016.05.040. http://www.sciencedirect.com/science/article/pii/S095070511630137X. New Avenues in Knowledge Bases for Natural Language Processing
5. Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D (eds) LREC. European Language Resources Association. http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf
6. Bagheri A, Saraee M, de Jong F (2013) Care more about customers: unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. Knowl-Based Syst 52(0):201–213
7. Brychcin T, Konkol M, Steinberger J (2014) Uwb: machine learning approach to aspect-based sentiment analysis. SemEval 2014:817
8. Burnap P, Williams ML (2015) Cyber hate speech on twitter: an application of machine classification and statistical modeling for policy and decision making. Policy Internet 7(2):223–242. https://doi.org/10.1002/poi3.85
9. De Marneffe MC, Manning CD (2008) The stanford typed dependencies representation. In: Coling 2008: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation, pp 1–8. Association for Computational Linguistics
10. Feldman R (2013) Techniques and applications for sentiment analysis. Commun ACM 56(4):82–89
11. Ghiassi M, Skinner J, Zimbra D (2013) Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. Expert Syst Appl 40(16):6266–6282
12. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224n Project Report, Stanford, pp 1–12
13. Hu M, Liu B (2004) Mining opinion features in customer reviews. In: Proceedings of the 19th National Conference on Artifical Intelligence, AAAI'04. AAAI Press, pp 755–760

14. Huang G, Huang GB, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48. https://doi.org/10.1016/j.neunet.2014.10.001. http://www.sciencedirect.com/science/article/pii/S0893608014002214

15. Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B Cybern 42(2):513–529

16. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: Nédellec C, Rouveirol C (eds) Machine learning: ECML-98, Lecture Notes in Computer Science, vol 1398. Springer, Berlin, pp 137–142

17. Kansal H, Toshniwal D (2014) Aspect based summarization of context dependent opinion words. Procedia Comput Sci 35(0):166–175. Knowledge-Based and Intelligent Information & Engineering Systems 18th Annual Conference, KES-2014 Gdynia, Poland, September 2014 Proceedings

18. Lek HH, Poo D (2013) Aspect-based twitter sentiment classification. In: 2013 IEEE 25th International Conference on Tools with Artificial Intelligence (ICTAI), pp 366–373

19. Li S, Zhou L, Li Y (2015) Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures. Inf Process Manag 51(1):58–67

20. Li Y, Qin Z, Xu W, Guo J (2015) A holistic model of mining product aspects and associated sentiments from online reviews. Multimed Tool Appl 74(23):10177

21. Liu B (2012) Sentiment analysis and opinion mining. Morgan & Claypool Publishers, San Rafael

22. Liu KL, Li WJ, Guo M Emoticon smoothed language models for twitter sentiment analysis. pp 1678–1684 (2012). Cited By (since 1996)1

23. Marrese-Taylor E, Velasquez JD, Bravo-Marquez F (2014) A novel deterministic approach for aspect-based opinion mining in tourism products reviews. Expert Syst Appl 41(17):7764–7775

24. Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41. https://doi.org/10.1145/219717.219748

25. Ou G, Chen W, Liu P, Wang T, Yang D, Lei K, Liu Y (2013) Aspect-specific polarity-aware summarization of online reviews. In: Wang J, Xiong H, Ishikawa Y, Xu J, Zhou J (eds) Web-Age Information Management, Lecture Notes in Computer Science, vol 7923. Springer, Berlin, pp 289–300

26. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing - volume 10, EMNLP '02. Association for Computational Linguistics, Stroudsburg, pp 79–86

27. Prabowo R, Thelwall M (2009) Sentiment analysis: a combined approach. J Inform 3(2):143–157. https://doi.org/10.1016/j.joi.2009.01.003. http://www.sciencedirect.com/science/article/pii/S1751157709000108

28. Sabbah T, Selamat A, Selamat MH, Ibrahim R, Fujita H (2016) Hybridized term-weighting method for dark web classification. Neurocomputing 173(3):1908–1926

29. Saif H, Fernandez M, He Y, Alani H (2013) Evaluation datasets for twitter sentiment analysis: a survey and a new dataset the sts-gold

30. Selamat A, Omatu S (2004) Web page feature selection and classification using neural networks. Inf Sci 158:69–88

31. Tripathy A, Agrawal A, Rath SK (2016) Classification of sentiment reviews using n-gram machine learning approach. Expert Syst Appl 57:117–126. https://doi.org/10.1016/j.eswa.2016.03.028. http://www.sciencedirect.com/science/article/pii/S095741741630118X

32. Vinodhini G, Chandrasekaran MR (2014) Opinion mining using principal component analysis based ensemble model for e-commerce application. CSI Trans on ICT 2(3):169–179

33. Vinodhini G, Chandrasekaran R (2016) A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. J King Saud University - Comput Inform Sci 28(1):2–12. https://doi.org/10.1016/j.jksuci.2014.03.024. http://www.sciencedirect.com/science/article/pii/S1319157815001020

34. Yu W, Zhuang F, He Q, Shi Z (2015) Learning deep representations via extreme learning machines. Neurocomputing 149, Part A:308–315. Advances in Neural Networks Advances in Extreme Learning Machines Selected papers from the Tenth International Symposium on Neural Networks (ISNN 2013) Selected articles from the International Symposium on Extreme Learning Machines (ELM 2013)

35. Zainuddin N, Selamat A (2014) Sentiment analysis using support vector machine. In: 2014 International Conference on Computer, Communications, and Control Technology (i4CT), pp 333–337

36. Zainuddin N, Selamat A, Ibrahim R (2016) Improving twitter aspect-based sentiment analysis using hybrid approach. Springer, Berlin, pp 151–160

37. Ziegelmayer D, Schrader R (2012) Sentiment polarity classification using statistical data compression models. In: 2012 IEEE 12Th International Conference on Data Mining Workshops, pp 731–738. https://doi.org/10.1109/ICDMW.2012.43

Nurulhuda Zainuddin is a PhD student at the Faculty of Computing, Universiti Teknologi Malaysia, Malaysia. She received a Bachelor of Science (Hons) Intelligent System from Universiti Teknologi MARA, Malaysia and a master's degree in Computer Science from Universiti Teknologi Malaysia, Malaysia. Her research interest focus on sentiment analysis and opinion mining, information retrievals and data mining.

Ali Selamat is a professor at Faculty of Computing, Universiti Teknologi Malaysia, Malaysia. He is currently a Chief Information Officer (CIO) and Director of Center for Information and Communication Technology, Universiti Teknologi Malaysia (UTM). He was the Dean of Research, Knowledge Economy Research Alliance, UTM. He is currently a Chair of IEEE Computer Society Malaysia Chapter. He is also the Editorial Boards of Knowledge Based Systems – Elsevier, International Journal of Information and Database Systems (IJIIDS), Inderscience Publications, Vietnam Journal of Computer Science, Springer Publications. He is the Program Chair of International Conference on Software Engineering Tools, Methodologies (SOMET), 2017 to be held in Kanazawa, Japan. He has been a visiting professor at Hradec-Kralove University, Czech Republic, Kuwait University, and few other universities in Japan. His research interests include cloud based software engineering, software agents, information retrievals, pattern recognitions, genetic algorithms, neural networks and soft-computing, knowledge management, key performance indicators.

Roliana Ibrahim is the Head of Department in Information Systems Department, Faculty of Computing, Universiti Teknologi Malaysia (UTM). She received her BSc (Honours) Computer Studies from Liverpool John Moores University, MSc Computer Science from Universiti Teknologi Malaysia and PhD in the field of Systems Engineering from Loughborough University. She has been an academic staff at Information Systems Department for since 1999 and was previously a Coordinator of BSc Computer Science (Bioinformatics) Program and Master of Information Technology (IT Management).