

Automatic classification of product reviews into interrogative and non-interrogative: Generating real time answer



Sheikh Muhammad Saqib ^{1,*}, Fazal Masud Kundi ¹, Shakeel Ahmad ², Tariq Naeem ¹

¹Institute of Computing and Information Technology, Gomal University, Dera Ismail Khan, Pakistan

²Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdul Aziz University (KAU), Jeddah, Saudi Arabia

ARTICLE INFO

Article history:

Received 12 March 2019

Received in revised form

3 June 2019

Accepted 4 June 2019

Keywords:

Sentiment classification

Preprocessing

Text mining

Sentiment orientation

ABSTRACT

Posted reviews on the relevant webpages about a product not only motivate the company to enhance quality but also it helps users to decide in favor of (or against) purchasing the product. These reviews are classified by different researchers through subjectivity based, entity based, or aspect based to find the polarity using the supervised or unsupervised technique. However, classification based on interrogatives and non-interrogatives is not handled yet. Datasets of interrogatives are analyzed as identifying Answer Seeking questions from Arabic tweets, question conveying and not conveying Information, Rhetorical Questions while here classifying the sentences into interrogatives and non-interrogatives is the preliminary step, which is a core contribution of proposed work. If detected questions are answered and moreover real time, it could not only motivate a user positively to buy the product but also users feel full duplex communication. In this work, we formulated this problem proposing linguistic and heuristic rules that automatically senses the interrogative and answer promptly based on the aforementioned aspect. If there is no aspect in an asked question, then LSI (Latent Semantic Indexing) generate answer using classified non-interrogatives. LSI is an efficient information retrieval algorithm, which finds the closest document to a given query. Experimental results using two publically available datasets show a precision of 95% and 96% which has 10% increased performance than alternatives machine learning methods Meta Filtered Classifier and Naive Bayes.

© 2019 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. Opinion extraction, sentiment mining, opinion mining, subjectivity analysis, emotion analysis, affect analysis, review mining, etc., are the different names and tasks of sentiment analysis. However, they are now all under the umbrella of sentiment analysis or opinion mining. While in industry, the term sentiment analysis is more commonly used, but in

academia both sentiment analysis and opinion mining are frequently employed.

As discussed above, pervasive real-life applications are only part of the reason why sentiment analysis is a popular research problem. It is also highly challenging as an NLP research topic and covers many novel sub problems. In NLP or in linguistic has little research before 2000 because at that time there was less amount of opinion text as digital form. After 2000, this field became most active research area in NLP with extensions of informational data mining and web mining. In fact, it has spread from computer science to management sciences (Liu, 2012; Archak et al., 2007; Chen and Xie, 2008). Useful information is known as data and lot or huge amount of data is known as big-data. Data mining means extraction of meaningful information from data. Text mining is the sub field of data mining. Major important data collection source is known as web mining. Opinion mining belongs to web content mining (Rubini and Chezian, 2014). For Sentiment classification work done on opinions are: - Subjectivity Classification, -Mapping Implicit Aspect,

* Corresponding Author.

Email Address: saqibsheikh4@gu.edu.pk (S. M. Saqib)

<https://doi.org/10.21833/ijaas.2019.08.004>

Corresponding author's ORCID profile:

<https://orcid.org/0000-0002-4647-1698>

2313-626X/© 2019 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

-Co-Reference Resolution Word, -Grouping Aspect into Category, -Dealing with Sarcastic Sentences. Some work has also been done on Comparative Opinions to generate Contrastive View Summary (Lerman and McDonald, 2009). Besides covering classifications of reviews from different dimensions, there is still need of classification of reviews with respect to interrogatives and non-interrogatives. In this regard, question conveying and not conveying Information (Zhao and Mei, 2013), identifying Answer Seeking questions from Arabic tweets (Hasanain et al., 2014), extraction of subjective/objective and questions and Rhetorical Questions (Hasanain et al., 2014; Liu and Jansen, 2015; Ranganath et al., 2016; Liu and Jansen, 2016), investigation of questions asked by Arab journalists (Hasanain et al., 2016), and detection of user intent behind asking the question (Kharche and Mante, 2017) have been studied. However, they all used the data readily available in the form of interrogatives. On the other hand, classifying the sentences in to interrogatives and non-interrogatives is the preliminary step.

Li et al. (2011a) coined the notion of Qtweets, the tweets which contain interrogative information that must be answered. For Qtweets, they detect a sentence which ends with question mark or a word starting from 5W1H-words. Resultantly, a sentence like “what a great job” if replaced as “what is a great job”, becomes interrogative. However, it is not an interrogative sentence in real sense. We not only cover these deficiencies but also improved this work. This study makes following key contributions:

- We propose heuristic rules for classification of subjective reviews into interrogatives and non-interrogatives.
- We propose a generation of real time answer from interrogative based on extracted aspects with user friendly interface.
- We propose a generation of real time answers using Latent Semantic Indexing (LSI) scores (without aspect).

2. Literature review

In order to attract customers, to motivate them into buying a product, and in order to get the feedback necessary for the assessment and improvement of the product, different companies have created opinion pages on their websites. Instead of going through the lengthy process of searching for the product, finding it and evaluating it, researchers have applied different techniques to separate, sift and distinguish the opinions in terms of the products. In order to separate the opinions, it is necessary to know about which particular product an opinion is being given. This process is termed ‘Entity Extraction’. Before the Entity Extraction can be carried out, it must be determined whether a given comment is an opinion or not. Objective opinions show factual information while the

subjective ones are users’ personal opinions (Liu, 2015); so, these can be considered for further processing because there may be sentiment in subjective opinion. Users express their views either in implicit way or explicit way: explicit opinions can be detected easily, while the implicit opinions are very hard to detect (Zhang and Liu, 2011; Greene and Resnik, 2009). Now to detect an entity from an opinion has been done using different methods. Authors Greene and Resnik (2009) have used two sets i.e. the set of seed entities Q and the set of candidate entities D to determine which of the entity in D belongs to C. For entity extraction (Pantel et al., 2009; Lee, 1999) have used a method distribution similarity by comparing the similarity of the surround words of each candidate entity with those of the seed entities; and then ranking the candidate entities based on the similarity values. Topic modeling has also been used for entity extraction. Topic modeling is an unsupervised learning method that assumes each document consists of a mixture of topics and each topic is a probability distribution over words (Liu, 2012). Latent Dirichlet Analysis LDA and Probabilistic Latent Semantic Analysis (PLSA) have been used for detection of topic from a document/documents (Blei et al., 2003; 2010; Griffiths and Steyvers, 2003). To search a required document from a huge amount of published articles is very time consuming and laborious work, topic modeling offers a computational tool to find relevant topics by capturing meaningful structure among the collections of documents (Wang et al., 2016). To obtain fine grained sentiment analysis, researchers have done work on aspects. Method defined by Long et al. (2010) used extracted nouns based on the frequency and information distance as aspects. Noun phrase that has sentiment bearing sentences can be considered as aspects (Blair-Goldensohn et al., 2008). Logic Programming, particularly Answer Set Programming (ASP), has been used to elegantly and efficiently implement the key components of syntax-based aspect extraction. Logic Programming provides a convenient and effective tool to encode and thus test knowledge needed to improve the aspect extraction methods, so that the researchers can focus on the identification and discovery of new knowledge to improve aspect extraction (Saqib et al., 2016b; Liu et al., 2013). A supervised learning algorithm has been used to extract aspects from an opinion of the product. The projected system implements aspect extraction using frequent item set at phrase level (Jeyapriya and Selvi, 2015; Ahmad et al., 2017). After the extraction of aspects, sentiment analysis can be done. Sentiment analysis means either opinion is positive or negative. This sentiment analysis can be done on all the opinions of all the products of a company; or on a particular product (entity extraction) (Batra and Rao, 2010; Engonopoulos et al., 2011); or on a particular feature (aspect extraction) of a product (Kirange et al., 2014; Brun et al., 2014; Pontiki et al., 2016; Alghunaim, 2015). To improve the precision and accuracy of each of the above mentioned techniques for

sentiment analysis, researchers have also done work on the detection of spam opinion (Saqib et al., 2018b; Li et al., 2011b; Ott et al., 2011; Teli and Biradar, 2014), co-reference resolution (Ding and Liu, 2010), detection of sense of ambiguated word (Saqib et al., 2018a), and aspects grouping (different words belonging to same aspects) (Saqib et al., 2019; Garcia-Pablos et al., 2014). After sentiment analysis on direct opinion, there is need to handle comparative sentence, interrogative sentences, etc. Work of Kwong and Yorke-Smith (2012) has detected the question-answer pairs from email threads to construct the summaries and Li and Zhao (2013) detected imperative sentence with interrogative mood. All the above-mentioned works are about the opinions that end-users post on the relevant webpages. These opinions are important for new users as they help them evaluate the product. But, at the same time, the questions they themselves ask about the product are even more significant than the general opinions of other people about the product. Users will be able to obtain the results through a sentiment analysis of the opinions. But it is even more desirable, useful and effective if users were to get answer to their questions automatically. For this purpose, interrogative sentences need to be separated from opinions.

3. Proposed framework

Customers in today's world are more concerned about the efficiency of the product they are going to buy. That is why they like to know about the product before they actually buy it through comments on webpage. Keeping these factors in mind, the authors realized that by looking at the progress and market value on the comments' page, the end user can ask any kind of question about the product. The other users can answer that question only when they visit that page. There has to be an automatic and, at the same time, effective system that is able to sense and understand an interrogative sentence, and to answer it on the spot.

3.1. Modules in proposed work

The proposed methodology consists of four modules: Classification Based on Interrogative and Non-Interrogative. Extraction of Aspects, Generation of Answers Based on Extracted Aspects and Answer through LSI (Latent Semantic Indexing) without Aspects as shown in Fig. 1.

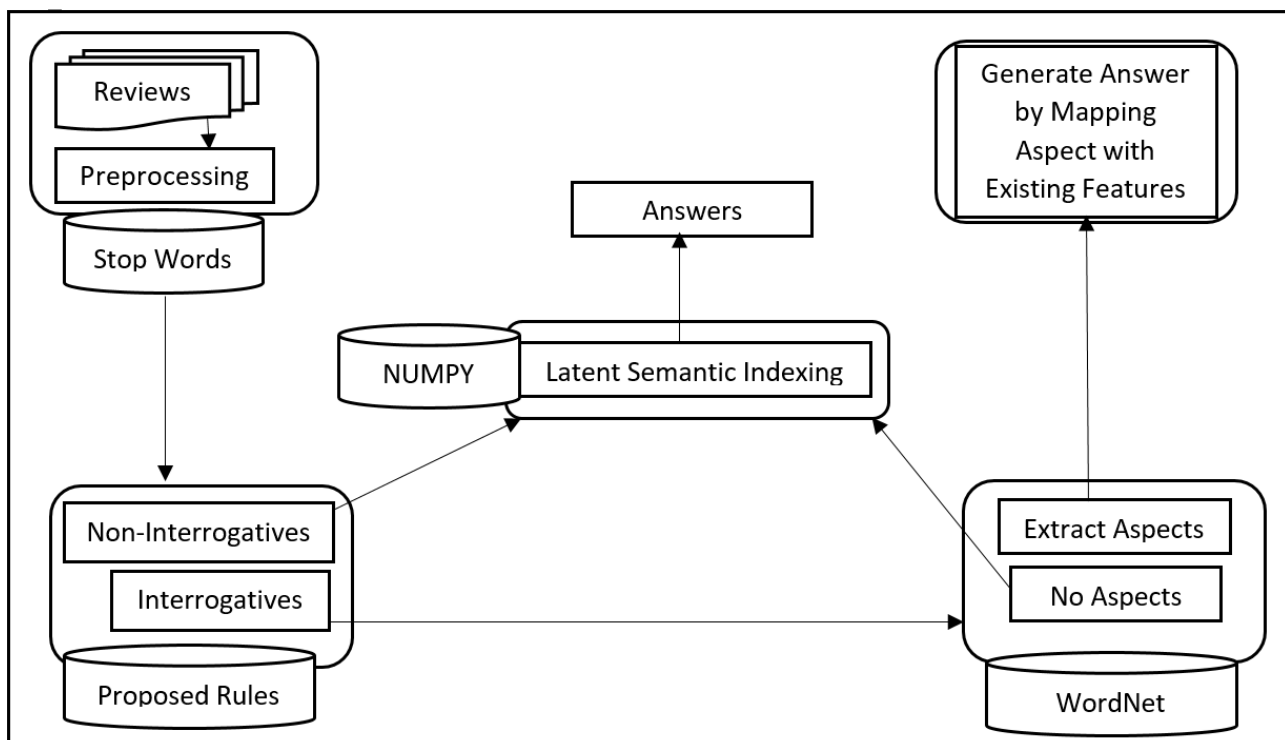


Fig. 1: Proposed framework

The following algorithm in Table 1 shows the entire process of the proposed methodology, where the 'getData' list contains the specifications of a product; the 'IsInterrogative' function will check whether the given comments are interrogative or not; 'getAspect' will determine the aspects from the interrogatives; and 'answerSummary' will generate an answer based on the extracted aspects.

Table 1: Algorithm for proposed work

Input Sentence as S
OutPut Answer Based on Aspects
Begin
If IsInterrogative (S) then Aspects= getAspect (S))
Summary =answerSummary(Aspects)
End if
Print Summary
End

All used functions i.e., IsInterrogative(), getAspect() and answerSummary() have been defined in Table 2, Table 3, and Table 4.

3.1.1. Classification based on interrogative and non-interrogative

These can be extracted through the following steps:

By "?" symbol

If a sentence ends with a "?" symbol, it means it is an interrogative sentence, so it is directly sent to the next step (i.e., Extract Aspect). The Rule-1 can be written as:

"?" ∈ Sentence => Interrogative (Rule-1)

By helping verb

Although interrogative sentences start with helping verbs, a list of helping verbs (HV) can be easily generated. If the starting word (SW) of a sentence belongs to the list of helping verbs (HV), it means it is an interrogative sentence as in Rule-2.

[SW] ∈ [HV] => Interrogative (Rule-2)

By W-family words

If the first word of the sentence belongs to the W-Family (What, Where, Why, When, Who, Whom, Whose), it means that the sentence is interrogative as defined in Rule-3.

[SW] ∈ [W – Family] => Interrogative (Rule-3)

But "What a great job", here, although the sentence starts with a W-family word, it is not an interrogative sentence. Then Rule-4 will be applied by finding second word (SecW) as HV.

[SW] ∈ [W – Family] and [SecW] ∈ [HV] => Interrogative (Rule-4)

Rule-4 is true on "How are you?" but false on "How this process will work?"; "How can I delete this file?", now Rule-5 will be considered as if any next word (ANW) belongs to HV, then it is considered as Interrogative.

[SW] ∈ [W – Family] and [ANW] ∈ [HV] => Interrogative (Rule-5)

The algorithm for detecting interrogatives is shown in Table 2 below where HV contains all helping verbs (do, did, has, have, could, would etc.).

3.1.2. Extraction of aspects

Aspects can be extracted using nouns and adjectives. They can be easily extracted using the tags of NLTK. After the removal of stop words i.e., "is, to, the, on, etc." there may be only nouns or adjectives in a single-line question. These nouns (NN) and adjective (JJ) are regarded as Aspects. The

algorithm for extracting aspects from interrogatives is shown in Table 3.

Table 2: Algorithm for interrogatives detection

Input Helping Verb List HV
Input W-Family Words IList
OutPut Classified Interrogatives in filter[]

Function IsInterrogative (Sentence as q)
if "?" in q:
filterQ.append (q)
else:
tk = wordTokenize(q)
if tk[0].lower() in HV:
filterQ.append (q)
else: if tk[0].lower() in IList and HV in tk:
filterQ.append ((q))
EndIf EndIf EndIf
End function

Table 3: Algorithm for aspects detection

Input Tokenize package of NLTK
Input Tag package of NLTK
Input StopWords
Output Aspects from given Sentence as q

Function getAspect (Sentence as q) tk = wordTokenize(q) aspects=""
for t in tk AND t not in StopWords AND if t in NN or JJ:
aspects = aspects+ t + " "
getAspect.append(aspects)
End if End for
End Function

3.1.3. Generation of answers based on extracted aspects

The aspects can be compared with the specifications of the product on the relevant webpage. The algorithm for generating answers from the extracted aspects is shown in Table 4.

Table 4: Algorithm for generating answers

Input Take specifications of a product using list getData=[]
Output Answer based on aspects

Function answerSummary (getAspects)
tk=wordTokenize(getAspects)
answer=""
for t in tk:
if t not in stopWords:
for g in getData:
if t in g: answer = answer + g + "LineBreak"
if answer=="": answerSummary.append("No Summary")
else: answerSummary.append(ans)
End if End if
End for End if End for
End Function

3.1.4. Answer through LSI (Latent semantic indexing) without aspects

Latent Semantic Indexing (LSI), proposed by Deerwester et al. (1990) in Blei et al. (2003), is an efficient information retrieval algorithm. Basically, LSI is a cosine similarity measure between the coordinates of a document vector and the coordinates of a query vector with the help of Singular Value Decomposition by using NUMPY package of python. If this value is 1, it means the document is 100% closer to the query, if it is 0.5, it means the document is 50% closer, and if it is 0.9, it means the document is 90% closer to the query

(Grossman and Frieder, 2012; Saqib et al., 2016a). If an interrogative has no aspects, then the LSI will determine the closest review for user satisfaction. The algorithm in Table 5 is used to generate answer for interrogative (with no answer), which will use question as query and remaining all comments as dataset.

Table 5: Algorithm of LSI

1.	Input: All Reviews and Query (Interrogative with No Answer)
2.	Tokenize All Reviews: $Token = Token(All\ Reviews)$
3.	Take the Union Set of Tokenized Reviews: $UnionT = Union(Token)$
4.	Make Frequency Matrix from $UnionTfMat = FrequencyMatrix(UnionT)$
5.	Make Query Matrix
6.	$qMat = QueryMatrix(Token(Query))$
7.	Decompose Frequency Matrix in U, S, V using SVD from USVT
8.	Determine V from VT
9.	Find UK, V_k and SK
10.	$UK =$ Extracting first two column of U
11.	$V_k =$ Extracting first two column of V
12.	$SK =$ Extracting first two column and row of S
13.	Each row of V relates to Coordinates of Document
14.	Find Coordinates of Query from $q = q^T * Uk * Sk^{-1}$
15.	First, we will find SK inverse from Step 10
16.	Second q transpose from Query Matrix from Step 8
17.	UK is already determined in Step 8
18.	Now, find $q = q^T * Uk * Sk^{-1}$
19.	q have coordinates of query
20.	Find dot product of q with each document coordinates (From 13)
21.	Sort dot product values in descending order
22.	Output Ranking of Reviews with respect to query
23.	Takes Reviews with score ≥ 0.99

4. Mathematical model of proposed work

Whole process for classification can be calculated from Eq-1 to Eq-7:

$$R = \bigcup_{x=1}^n R_x \quad (1)$$

$$T(x) = \bigcup_{i=1}^n T_i \quad (2)$$

$$FT(x) = \bigcup_{i=1}^n \{T_i, \text{ if } T_i \notin SW\} \quad (3)$$

where $x = 1, 2, 3, \dots, n$, SW means stop words, R represents the total number of reviews, $T(x)$ represents the tokens of the x^{th} review, and $FT(x)$ represents the filtered tokens of the x^{th} review.

$$Cla(x) = H_{Interrogative}(FT(x)) \quad (4)$$

$$H_{Interrogative}(x) = \begin{cases} I_x, & \text{if } (w_1 \text{ or } w_n) \in (LR \text{ or } HR) \\ \sim I_x, & \text{else} \end{cases} \quad (5)$$

where $x = 1, 2, 3, \dots, n$, $Cla(x)$ will classify the x^{th} review as being interrogative or not. w_1 and w_n are the first and last words of a review, I_x means the x^{th} review is interrogative, and $\sim I_x$ means the x^{th} review is not interrogative. LR and HR are the linguistic rules and heuristic rules, respectively. The resultant interrogatives and non-interrogatives can be grouped as follows:

$$Cla^I = \bigcup_{i=1}^n I_i \quad (6)$$

$$Cla^{\sim I} = \bigcup_{i=1}^n \sim I_i \quad (7)$$

5. Experimental results

We have tested this model on Dataset-1 and Dataset-2. Dataset-1 1031 questions (as interrogatives) and 1031 answers (as Non-Interrogatives) downloaded from <http://www.cs.cmu.edu/ark/QA-data/>. This page provides a link to a corpus of Wikipedia articles, manually-generated factoid questions from them, and manually generated answers to these questions, for use in academic research. These data were collected by Noah Smith, Michael Heilman, Rebecca Hwa, Shay Cohen, Kevin Gimpel, and many students at Carnegie Mellon University and the University of Pittsburgh between 2008 and 2010. Version 1.2 released August 23, 2013 has manually-generated factoid question/answer pairs with difficulty ratings from Wikipedia articles. This Dataset includes articles, questions, and answers. Dataset-2 based on product reviews of Samsung Galaxy J7 from URL (http://www.gsmarena.com/samsung_galaxy_j7-7185.php visit on Date 31 August 2017) to check the performance of proposed work. This page has more than 5000 reviews from different users of the product. These datasets are arranged as Interrogatives and Non-Interrogatives manually to create 'arrf' file for machine learning algorithms and text file for propose work.

5.1. Statistical measures

A confusion matrix is formed from the four outcomes produced as a result of binary classification. A binary classifier predicts all data instances of a test dataset as either positive or negative. This classification (or prediction) produces four outcomes—true positive, true negative, false positive and false negative. True positive (TP): correct positive prediction, False positive (FP): incorrect positive prediction, True negative (TN): correct negative prediction, False negative (FN): incorrect negative prediction. Various measures can be derived from a confusion matrix as shown in Table 6.

We compare our proposed method with the following machine learning methods for sentence classification:

- Method-1 (Naive Bayes): In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. The Naive Bayesian classifier is based on Bayes' theorem with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

- ii) Method-2 (Meta Filtered Classifier): This Class is used for running an arbitrary classifier on data that has been passed through an arbitrary filter. Similar to classifier, the structure of the filter is based exclusively on the training data and test instances will be processed by the filter without changing their structure (Vijayarani and Muthulakshmi, 2013).
- iii) Proposed Model: As shown in Table 1, Table 2, Table 3, and Table 4, the proposed method achieved more robust and more accurate results by proposing five rules to classify sentences into interrogatives and non-interrogatives. After

experimental results on Dataset-1, all answers have been considering as non-interrogatives because no rule related to question mistakenly consider as interrogatives. All questions end with “?” while we have omitted such symbol and then trying to detect questions because sometimes user is not put “?” at the end on review page. Hence, this agent had average precision of 95% and accuracy 94% on Dataset-1 and average precision of 96% and accuracy 94% on Dataset-2 in detecting interrogatives (I) and non-interrogatives (NI) as shown in Table 7.

Table 6: Statistical measures

Measure	Derivations
Sensitivity or True Positive Rate	(TPR) $TPR = TP / (TP + FN)$
Specificity or True Negative Rate	(TNR) $SPC = TN / (FP + TN)$
Precision or Positive Predictive Value	(PPV) $PPV = TP / (TP + FP)$
Negative Predictive Value	$NPV = TN / (TN + FN)$
False Positive Rate	$FPR = FP / (FP + TN)$
False Discovery Rate	$FDR = FP / (FP + TP)$
False Negative Rate	$FNR = FN / (FN + TP)$
Accuracy	$ACC = (TP + TN) / (P + N)$
F1 Score	$F1 = 2TP / (2TP + FP + FN)$

Table 7: Comparison of proposed classification technique with machine learning methods

	Dataset-1									Class
	TPR	SPC	PPV	NPV	FPR	FDR	FNR	ACC	F1	
Naive	0.8182	0.8889	0.9	0.8	0.1111	0.1	0.1818	0.85	0.8571	Interrogative
Bayes	0.8889	0.8182	0.8	0.9	0.1818	0.2	0.1111	0.85	0.8421	Non-Interrogative
Meta	0.85355	0.85355	0.85	0.85	0.14645	0.15	0.14645	0.85	0.8496	Average
Filtered	1	0.7407	0.65	1	0.2593	0.35	0	0.825	0.7879	Interrogative
Classifier	0.7407	1	1	0.65	0	0	0.2593	0.825	0.8511	Non-Interrogative
	0.87035	0.87035	0.825	0.825	0.12965	0.175	0.12965	0.825	0.8195	Average
Proposed	0.8889	1	1	0.9091	0	0	0.1111	0.9474	0.9412	Interrogative
Method	1	0.8889	0.9091	1	0.1111	0.0909	0	0.9474	0.9524	Non-Interrogative
	0.94445	0.94445	0.95455	0.95455	0.05555	0.04545	0.05555	0.9474	0.9468	Average
	Dataset-2									Class
	TPR	SPC	PPV	NPV	FPR	FDR	FNR	ACC	F1	
Naive	0.8235	1	1	0.4	0	0	0.1765	0.8421	0.9032	Interrogative
Bayes	1	0.8235	0.4	1	0.1765	0.6	0	0.8421	0.5714	Non-Interrogative
Meta	0.91175	0.91175	0.7	0.7	0.08825	0.3	0.08825	0.8421	0.7373	Average
Filtered	0.7368	0	1	0	0	0	0.2632	0.7368	0.8485	Interrogative
Classifier	0	0.7368	0	1	0.2632	1	0	0.7368	0	Non-Interrogative
	0.3684	0.3684	0.5	0.5	0.1316	0.5	0.1316	0.7368	0.42425	Average
Proposed	1	0.8333	0.9286	1	0.1667	0.0714	0	0.9474	0.963	Interrogative
Method	0.8333	1	1	0.9286	0	0	0.1667	0.9474	0.9091	Non-Interrogative
	0.91665	0.91665	0.9643	0.9643	0.08335	0.0357	0.08335	0.9474	0.93605	Average

6. Conclusion and future work

This work proposed classification of product reviews based on interrogatives and non-interrogative. We obtained classification results with improved precision (0.95) when compared to the alternative machine learning methods as shown in Fig. 2.

The proposed method is quite generalized, and it can classify reviews of any product, android applications etc.

The proposed method does feature several limitations that may be borne in mind when interpreting the findings, which are discussed in below using different samples of questions. Some are shown in Table 8.

In the above samples, S. No. 1 and S. No. 2 are not interrogatives because they do not start with either a W-Family word or a helping verb. For S. No. 3 (It is

dual sim?); if there had been no interrogation mark (?), then the agent would not have detected them as questions. S. No. 4 (what bands it supports) is actually a question, but the agent did not detect it due to the fact that even though it starts with a W-Family word, yet there is no helping verb in the entire sentence. So, the rule could not succeed here and that is why it was considered as “Not Interrogative”. If the user had written it as, “What bands it can support”, then the agent would have been able to detect it even without the question mark, because of the W-Family word, “what”, and the helping verb, “can”. All the possible patterns and modes of questions using W-Family words or the ones using helping verbs in the beginning were incorporated. Utmost care was taken to ensure that all such sentence templates were covered. The authors were confident that this sample was sufficient to serve as substantial evidence to support

the arguments established in this study. In Table 9, there are some sample of detected interrogatives whose answers has generated on aspects. Here, the

answer to S. No 4 was generated through LSI score in a Table 10, because there is no aspect.

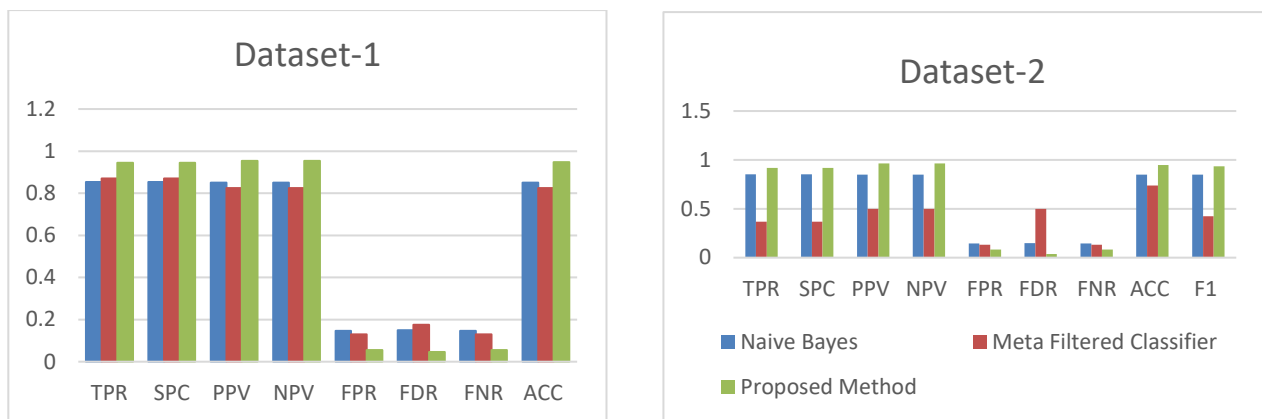


Fig. 2: Comparison of proposed measures with machine learning methods

Table 8: Detected orientation of sentences by proposed work

S.No.	Sentences	Detected Orientation	Result	Rules
1	What a great job	Non-Interrogative	True	Rule-4
2	It does not have its own music player	Non-Interrogative	True	No
3	It is dual sim?	Interrogative	True	Rule-1
4	What bands it supports	Non-Interrogative	False	No
5	Is it Dual SIM	Interrogative	True	Rule-2
6	What are settings	Interrogative	True	Rule-3
7	When it was Announced	Interrogative	True	Rule-5
8	J7 is working properly?	Interrogative	True	Rule-1

Table 9: Answers generated by agent based on detected questions

S. No.	Interrogatives	Extracted Aspects (NN, JJ)	Answers on Aspects.
1	It is dual sim?	dual	SIM 1 (dual-SIM model only)
2	what is Internal memory	Internal	Internal: 16 GB, 1.5 GB RAM:
3	what is the Speed of this mobile	Speed	Speed: HSPA 21.1/5.76 Mbps
4	J7 is working properly?	No Aspects	No Summary

Table 10: Closest review to interrogative with no aspects

Question	J7 is working properly?	LSI-Score
Answer-1	mobile data network is not working properly	0.999999262
Answer-2	J7 is the Best Smartphone. Superbbbb...	0.999898212649
Answer-3	worst performing smartphone till now.....low ram,	0.998200989159
Answer-4	worst mobile, sensitive touch display	0.982944211896

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Ahmad S, Saqib SM, Almagrabi AO, and Alotaibi FM (2017). LSI based search technique: Using extracted keywords and key-sentences. VAWKUM Transactions on Computer Sciences, 14(2): 1-8.
<https://doi.org/10.21015/vtcs.v14i2.471>
- Alghunaim A (2015). A vector space approach for aspect-based sentiment analysis. Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, USA.
<https://doi.org/10.3115/v1/W15-1516>
- Archak N, Ghose A, and Ipeirotis PG (2007). Show me the money: Deriving the pricing power of product features by mining consumer reviews. In the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Jose, USA: 56-65.
<https://doi.org/10.1145/1281192.1281202>

- Batra S and Rao D (2010). Entity based sentiment analysis on twitter. Science, 9(4): 1-12.
- Blair-Goldensohn S, Hannan K, McDonald R, Neylon T, Reis G, and Reynar J (2008). Building a sentiment summarizer for local service reviews. In the NLP Challenges in the Information Explosion Era, Beijing, China: 1-10.
- Blei D, Carin L, and Dunson D (2010). Probabilistic topic models: A focus on graphical model design and applications to document and image analysis. IEEE Signal Processing Magazine, 27(6): 55-65.
<https://doi.org/10.1109/MSP.2010.938079>
PMid:25104898 PMCID:PMC4122269
- Blei DM, Ng AY, and Jordan MI (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3: 993-1022.
- Brun C, Popa DN, and Roux C (2014). XRCE: Hybrid classification for aspect-based sentiment analysis. In the 8th International Workshop on Semantic Evaluation, Dublin, Ireland: 838-842.
<https://doi.org/10.3115/v1/S14-2149>
- Chen Y and Xie J (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. Management Science, 54(3): 477-491.
<https://doi.org/10.1287/mnsc.1070.0810>
- Deerwester S, Dumais ST, Furnas GW, Landauer TK, and Harshman R (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science,

- 41(6): 391-407.
[https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-AS11>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-AS11>3.0.CO;2-9)
- Ding X and Liu B (2010). Resolving object and attribute coreference in opinion mining. In the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, Beijing, China: 268-276.
- Engonopoulos N, Lazaridou A, Paliouras G, and Chandrinou K (2011). ELS: A word-level method for entity-level sentiment analysis. In the International Conference on Web Intelligence, Mining and Semantics, ACM, Sogndal, Norway: 1-9.
- Greene S and Resnik P (2009). More than words: Syntactic packaging and implicit sentiment. In the Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Boulder, Colorado, USA: 503-511.
<https://doi.org/10.3115/1620754.1620827>
- Griffiths TL and Steyvers M (2003). Prediction and semantic association. In: Sahani M and Linden JF (Eds.), Advances in neural information processing systems: 11-18. MIT Press, Cambridge, USA.
- Grossman DA and Frieder O (2012). Information retrieval: Algorithms and heuristics. Springer Science and Business Media, Berlin, Germany.
- Hasanain M, Bagdouri M, Elsayed T, and Oard DW (2016). What questions do journalists ask on twitter?. In the 10th International AAAI Conference on Web and Social Media, Association for the Advancement of Artificial Intelligence (AAAI), Cologne, Germany: 127-134.
- Hasanain M, Elsayed T, and Magdy W (2014). Identification of answer-seeking questions in Arabic microblogs. In the 23rd ACM International Conference on Conference on Information and Knowledge Management, ACM, Shanghai, China: 1839-1842.
<https://doi.org/10.1145/2661829.2661959>
- Jeyapriya A and Selvi CK (2015). Extracting aspects and mining opinions in product reviews using supervised learning algorithm. In the 2nd International Conference on Electronics and Communication Systems, IEEE, Coimbatore, India: 548-552.
<https://doi.org/10.1109/ECS.2015.7124967>
- Kharche SE and Mante RV (2017). An efficient way of answering the questions asked on social sites by understanding user intent. In the 2017 International Conference on Recent Trends in Electrical, Electronics and Computing Technologies, IEEE, Warangal, India: 159-163.
<https://doi.org/10.1109/ICRTEECT.2017.22>
- Kirange DK, Deshmukh RR, and Kirange MDK (2014). Aspect based sentiment analysis SemEval-2014 task 4. Asian Journal of Computer Science {and} Information Technology, 4(8): 72-75.
<https://doi.org/10.15520/ajcsit.v4i8.9>
- Kwong H and Yorke-Smith N (2012). Detection of imperative and declarative question-answer pairs in email conversations. AI Communications, 25(4): 271-283.
- Lee L (1999). Measures of distributional similarity. In the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, Association for Computational Linguistics, College Park, Maryland, USA: 25-32.
<https://doi.org/10.3115/1034678.1034693>
PMCID:PMC1232645
- Lerman K and McDonald R (2009). Contrastive summarization: An experiment with consumer reviews. In the Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Boulder, USA: 113-116.
<https://doi.org/10.3115/1620853.1620886>
- Li B, Si X, Lyu MR, King I, and Chang EY (2011a). Question identification on twitter. In the 20th ACM International Conference on Information and Knowledge Management, ACM, Glasgow, Scotland, UK: 2477-2480.
<https://doi.org/10.1145/2063576.2063996>
PMCID:PMC3116284
- Li FH, Huang M, Yang Y, and Zhu X (2011b). Learning to identify review spam. In the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain: 2488-2493.
- Li P and Zhao H (2013). NLP-oriented study on the imperative sentence with interrogative mood. In the 2013 International Conference on Asian Language Processing, IEEE, Urumqi, China: 54-56.
<https://doi.org/10.1109/IALP.2013.18>
- Liu B (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1): 1-167.
<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Liu B (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, Cambridge, UK.
<https://doi.org/10.1017/CBO9781139084789>
- Liu Q, Gao Z, Liu B, and Zhang Y (2013). A logic programming approach to aspect extraction in opinion mining. In the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies, IEEE, Atlanta, USA, 1: 276-283.
<https://doi.org/10.1109/WI-IAT.2013.40>
- Liu Z and Jansen BJ (2015). Subjective versus objective questions: Perception of question subjectivity in social Q&A. In the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Springer, Washington, USA: 131-140.
https://doi.org/10.1007/978-3-319-16268-3_14
- Liu Z and Jansen BJ (2016). Understanding and predicting question subjectivity in social question and answering. IEEE Transactions on Computational Social Systems, 3(1): 32-41.
<https://doi.org/10.1109/TCSS.2016.2564400>
- Long C, Zhang J, and Zhut X (2010). A review selection approach for accurate feature rating estimation. In the 23rd International Conference on Computational Linguistics: Posters, Association for Computational Linguistics, Beijing, China: 766-774.
- Ott M, Choi Y, Cardie C, and Hancock JT (2011). Finding deceptive opinion spam by any stretch of the imagination. In the 49th Annual Meeting of the Association for Computational Linguistics: Human language Technologies, Association for Computational Linguistics, Portland, Oregon, USA, 1: 309-319.
- Pablos AG, Cuadros M, and Rigau G (2014). V3: Unsupervised generation of domain aspect terms for aspect based sentiment analysis. In the 8th International Workshop on Semantic Evaluation, Dublin, Ireland: 833-837.
- Pantel P, Crestan E, Borkovsky A, Popescu AM, and Vyas V (2009). Web-scale distributional similarity and entity set expansion. In the 2009 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2: 938-947.
<https://doi.org/10.3115/1699571.1699635>
- Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, Mohammad AS, and Hoste V (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In the 10th International Workshop on Semantic Evaluation, Association for Computational Linguistics, San Diego, California, USA: 19-30.
- Ranganath S, Hu X, Tang J, Wang S, and Liu H (2016). Identifying rhetorical questions in social media. In the 10th International AAAI Conference on Web and Social Media, AAAI, Cologne, Germany: 667-670.
- Rubini R and Chezian RM (2014). An Analysis on search engines: Techniques and tools. International Journal of Advanced

- Research in Computer and Communication Engineering, 3(9): 7860-7864.
- Saqib SM, Ahmad S, Syed AH, Naeem T, Alotaibi FM (2019). Grouping of aspects into relevant category based on WordNet definitions. International Journal of Computer Science and Network Security, 19(2): 113-119.
- Saqib SM, Kundi FM, and Ahmad S (2018a). Unsupervised learning method for sorting positive and negative reviews using LSI (latent semantic indexing) with automatic generated queries. International Journal of Computer Science and Network Security, 18(1): 56-62.
- Saqib SM, Kundi FM, Syed AH, and Ahmad S (2018b). Semi supervised method for detection of ambiguous word and creation of sense: Using WordNet. International Journal of Advanced Computer Science and Applications, 9(11): 353-359.
<https://doi.org/10.14569/IJACSA.2018.091149>
- Saqib SM, Mahmood K, and Naeem T (2016a). Comparison of LSI algorithms without and with pre-processing: using text document based search. ACCENTS Transactions on Information Security, 1(4): 44-51.
- Saqib SM, Naeem T, and Mahmood K (2016b). Extraction of key/ title/ aspect words from document using WordNet. ACCENTS Transactions on Information Security, 1(3): 36-43.
- Teli S and Biradar S (2014). Effective spam detection method for email. In the International Conference on Advances in Engineering and Technology, Singapore: 68-72.
- Vijayarani S and Muthulakshmi M (2013). Comparative study on classification meta algorithms. International Journal of Innovative Research in Computer and Communication Engineering, 1(8): 1768-1774.
- Wang SH, Ding Y, Zhao W, Huang YH, Perkins R, Zou W, and Chen JJ (2016). Text mining for identifying topics in the literatures about adolescent substance use and depression. BMC Public Health, 16(1): 279-286.
<https://doi.org/10.1186/s12889-016-2932-1>
PMid:26993983 PMCID:PMC4799597
- Zhang L and Liu B (2011). Identifying noun product features that imply opinions. In the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Association for Computational Linguistics, Portland, Oregon, USA, 2: 575-580.
- Zhao Z and Mei Q (2013). Questions about questions: An empirical analysis of information needs on twitter. In the 22nd International Conference on World Wide Web, ACM, Rio de Janeiro, Brazil: 1545-1556.
<https://doi.org/10.1145/2488388.2488523>
PMCID:PMC3971240