Inferring the Products Realistic Feature Through Data From Users Views In Socialmedia

Sumithra M, Asha Abraham, Gracia Nissi

Abstract*: In current trends, online life becomes as an inevitable option, feasible source to remove extensive scale, heterogeneous item includes in a period and cost-proficient way. One of the difficulties of using social media data is to educate people with availability of item choices along with related information, for example, mockery, which represents 22.75% of web based data and can possibly make prediction in the predictive models that gain from such information sources. For instance, if a client says "I simply love holding up throughout the day while this tune downloads," a feature extraction model may mistakenly relate a positive estimation of "adoration" to the mobile phone's capacity to download. While conventional content mining strategies are intended to deal with all around framed content where item includes are gathered from the mix of words, these devices would neglect to process these social messages that incorporate understood implicit information conveyed through the data. In this paper, we propose a technique that empowers users to use understood social media data by making an interpretation of each verifiable message into its proportional express structure, utilizing the word simultaneousness organize as a coherence network of word (coward). A case study of Twitter messages that talk about Smartphone highlights is utilized to approve the proposed technique. The outcomes from the analysis not just demonstrate that the proposed strategy improves the interpretability of verifiable messages, yet additionally reveals insight into potential applications in the various fields where this work could be broadened.

Index Terms: Marketing, Segmentation, Technology and Buying Behaviour.

I. INTRODUCTION

The rigorous competition in the market space drives designers to create products that better satisfy the majority of customers in a resource efficient manner Oftentimes, it is crucial that designers are familiar with target customers' needs and preferences, in order to incorporate preferable features and remove weak elements from a design artifact. Recently, the literature has shown that information generated by social media users could prove critical to product designers in learning relevant preferences toward products/product features[1–6].The literature in various fields of study has shown successful applications that rely on information extracted from large-scale social media data, such as mining healthcare- related information for disease prediction [7–9], detecting earth- quake warnings and emergence needsdue to natural disasters [10,11], and predicting financial market movement [12,13].In the design informatics domain, despite the traditional methods that extract customers preferences from online product reviews, recent findings have illustrated that social networkscould also serve as a viable source of information for mining customers' opinions toward products/product features, due to its fast publication, wide range of users, accessibility, and heterogeneity of contents that provides an opportunity for customers to express opinions about products outside the review sites [2]. A data-driven methodology has been proposed to automatically discover notable product features mentioned in social networks [5]. Later, such notable product feature information is incorporated into a decision support framework that helps designers to develop next- generation products [2]. Furthermore, large-scale social media data have been established as a viable platform to automatically discover innovative users in social networks [1,4]. Such innovative users could prove critical to product design and development as they help designers to discover relevant product feature preferences months or even years before they are desired by general customers.

Implicit speech is a form of language usage in which the actual meaning is intended to be comprehended, but not directly stated. A majority manifestation of implicit speech includes sarcasm, which has become not only abundant, but also a norm in social networks. Maynard and Greenwood found that roughly 22.75% of social media data are sarcastic [14]. While it is evident that knowledge extracted from social media data is useful to product designers, the applicability of such data pertains to the portion expressed in explicit forms, due to the limitation of the underlying natural language processing algorithms that assume the explicit, well-formed textual input. As a result, implicit information would be either treated as noises or misinterpreted, resulting in inaccurate recommendation of product design decision support systems that process the information from large-scale social media data. Hence, the ability to automatically understand and correctly interpret such implicit information in social networks would not only reduce the errors caused, but also allow the methodologies to make use of additional implicit data that would have traditionally been disregarded

Revised Manuscript Received on December 22, 2018.

Sumithra M, Department of Computer Science and Engineering, VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology,Avadi, Chennai, India.
Asha Abraham, Department of Computer Science and Engineering, VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology,Avadi, Chennai, India.
Gracia Nissi S., Department of Computer Science and Engineering, Kings Engineering College, Chennai, India.

.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

due to being treated asnoise.
If implicit social media messages remain untreated, two problems could occur:
Many data mining algorithms are extraction based that would classify a social media message whether it is useful or not. Such methods would disregard such implicit data where explicit knowledge could not be extracted, resulting in low utilization of usefuldata.
Sarcasticsocialmediamessagesmayeitherexaggeratetheoriginalmeaning.Thetraditionaltextminingtechniquesareincapableofcorrectly interpreting the true meaning of these untreated social mediamessages.
Regardless of all the useful applications that emerge from social media data, being able to automatically explicate the implicit social media data would not only increase the performance of the existing natural language processing techniques, but would also enable discovery of real important product features that exist in the implicitdata.
Processing social media data has been one of the biggest challenges for researchers. Traditional natural language processing techniques that have been shown to work well on traditional documents are reported to fail or underperform when applied on social media data, whose natures differ from traditional documents in the following ways:
Social media data are high-dimensional, but sparse.
Social media data is noisy.
The existing attempts to interpret the semantic meaning behind implicit social media and relevant kinds of data (i.e., product reviews) include machine learning based implicit sentence detection algorithms proposed by Tsur and coworkers [16,17]. How- ever, their methods only identify whether a piece of textual information is sarcastic or not. The work presented in this paper extends the previous literature by further extracting true meaning from social media messages whose context related to products/ product features are implicit.

This paper presents a mathematical model based on the heterogeneous coword network patterns in order to translate implicit context toward a particular product or product feature into the explicit equivalence. A coword network (or word co-occurrence network) is a graph where each node represents a unique word, andanundirectededgerepresentsthefrequencyofco-occurrence ofthetwowords.Inthiswork,thenetworkisaugmentedtoincorporatepartsofspeechintoeachword.Theintuitionbehindusing the coword network is that even though a message may be implicit, the similar combination of the words may have been usedbyotherusers whoexpresstheirmessagesmoreexplicitly.
For example, given an implicit message “wow I have to squinttoreadthisonthescreen,” other usersmay haveusedthetermssquintandscreeninamoreexplicitcontext such as “Don’t make me squint @user - your mobile banner needs work on my tiny screen iPhone 5S.” If the combination of the words squint and screen occurs in the messagesthatcontainthewordtinyfrequentlyenough,thenthesys - tem would be able to relate the original message to a more explicit set of terms. Particularly, the system would be able to interpretthattheuserthinksthatthescreenfeatureofthisparticular product issmall.Specifically, this paper has the following main contributions:
The authors adopt the usage of the coword network in a product design context. The coword network has shown to be useful in multiple semantic extraction applications in information retrieval literature [7,18]. To the best of our knowledge, this technique has first been used in the design literature.
The authors propose a probabilistic mathematical model in order to map implicit product-related information in social media data into the equivalent explicitcontext.
The authors illustrate the efficacy of the proposed methodology using a case study of real world smartphone data and Twitterdata.

RELATEDWORKS

While the use of implicit language such as indirect speech and sarcasm has been well explored in multiple psycholinguistic studies [19–21], automatic semantic interpretation of implicit information in social networks is still in an infancy stage. This section first surveys the use of social media data pertaining to the product design applications and then discusses the existing natural language processing techniques that have been used to extractsemantics from social mediadata.
Applications of Large-Scale Social Media Datain Product Design Domain. Knowledge extracted from product- related, user-generated information has proved valuable in product design applications. Archak et al. proposed a set of algorithms, both fully automated and semi-automated, to extract opinionated product features from online reviews. The extracted information was successfully used to predict product demand [22]. While their findings were promising, the algorithms were applied on online product reviews whose nature is different from social media data, in terms of noise, amount of indirect language (i.e., sarcasm), and language creativity that do not conform to the standard English grammar. This research primarily aims to interpret semantics of a subset of social media data whose language is presented with sarcasm that traditional natural language processing techniques would fail to handle effectively. Social media is characterized as a major source for product design and development. Asur and Huberman were able to use Twitter data collected during a 3-month period to predict the demand of theater movies[23]. They claimed that the prediction results are more accurate than those of the Hollywood Stock Exchange. Their study also found that sentiments in tweets can improve the prediction after a movie has been released. Tuarob and Tucker found that social media data could be a potential data source for extracting user preferences toward particular products or product features [2,5]. In a later work, they exhibited an approach for programmed disclosure of creative clients (otherwise known as. lead clients) in online networks, utilizing a lot of scientific models to remove idle highlights (item includes not yet executed in the market space), at that point distinguish lead clients dependent on the volume of inventive highlights that they express in web based life [1,4].

Lim and Tucker proposed a Bayesian-based factual testing calculation that distinguishes item highlight related catchphrases from web based life information, without human-named preparing information [6]. As of late, Stone and Choi displayed a representation apparatus which enables originators to extricate valuable bits of knowledge from online item surveys [24].Since all the above techniques rely on the assumption that social media data are explicit, these techniques would fail to correctly process implicit social media messages which could result in error or inaccurate results. With these emerging product design applications that rely on social media as a knowledge source, it is crucial that the algorithms behind these applications are able to correctly interpret the true meaning of the data.

Social media holds sentiments expressed by its users (primarily in the form of textual data). Sentiment analysis is used to linguistics to identify and extract subjective information in social media. It involves natural language processing, text analysis, and computational. Thelwall et al. found that important events lead to increases in average negative sentiment strength in tweets during the same period [30]. The authors concluded that the negtive sentiment may be the key to popular trends in Twitter. Kucuktunc et al. studied the influence of several factors such as gender, age, education level, discussion topic, and time of day on sentiment variation in Yahoo! Answers [31]. Their findings shed light toward an application on attitude prediction in online question-answering forums. Weber et al. proposed a machine learning based algorithm to mine tips, short, self-contained, concise texts describing nonobvious advice [32]. Lim et al. applied unsupervised sentiment analysis in social media to identify the patient's potential symptoms and latent infectious diseases [9]. Sentiment of each short text is extracted and used as part of the features. Even though sentiment analysis could prove to be useful when designers would like to know how customers feel about a particular product or product feature, most sentiment extraction techniques only output sentiment level in two dimension (i.e., positive and negative). Hence, more advanced techniques are needed in order to narrow down what actually the customers want to say.

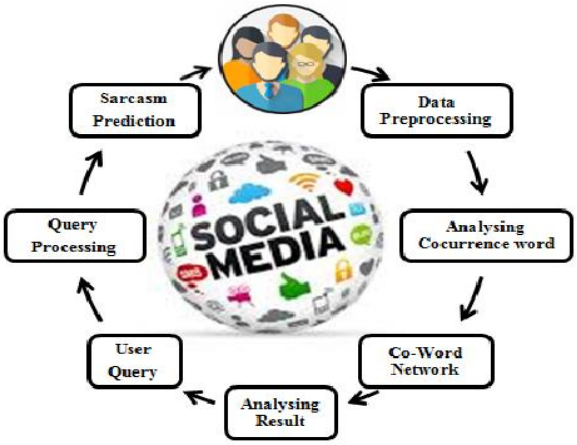
Besides sentiment analysis, multiple studies have foundthattopical analysis could be useful when dealing with noisy textual data such as social media. Even though social media is high in noise due to the heterogeneity of the writing styles, formality, and creativity, such noise bears undiscovered wisdom of the crowd. Paul and Dredze utilized a modified latent Dirichletallocation[33] model to identify 15 ailments along with descriptions and symptoms in Twitter data [34,35]. Tuarob et al. proposed a methodology for discovering health-related content in social media data by quantifying topical similarity between documents as a feature type [7,8]. Furthermore, a number of studies have devoted to usingtopicalmodelstodetectemergingtrendsinsocialnetworksf rom large-scale social media data, such as customer demands, notable product features, and innovative product ideas [1,2,39]. The techniques mentioned earlier rely on explicit content of social media data and would likely fail or not produce correct results when applied on documents whose meanings are implicit.

Implicit document processing has posed challenges to computational linguists. Researchers have studied on the nature of implicit uses of language; however, none have successfully developed a computational model to translate implicit content into the equivalent explicit form. In dealing with implicit context in social media data, multiple algorithms have been proposed to detect the presence of implicit content in social media [16,40,41]; To the best of our knowledge, we are the first to explore the problem of identifying explicit customer preferences toward a product/product feature from large- scale social mediadata.

II. METHODOLOGY

Themethodproposedinthispapermineslanguageusagesinthe form of word co-occurrence patterns, in order to map implicitcontext commonly found in social media data to equivalent explicit ones. First, social media data are collected and preprocessed (Sec. 3.1). The textual content is then fed to the indexer in order to generate the coword network (Sec. 3.2). After framing the network and labeled user can submit an implicit message in the form of query.The processed query return ranked list of classifieds based on the phrase.The system might be a human designer, or an automated program that mines product- related information from social media messages

Figure1-Proposedmethodology.



.

A practical usage of the proposed implicit message inference system would be to aid designers in synthesizing product features, mined from customers' feedback in large-scale social media data, into the next generation products. A framework was presented in Ref. [2], where designers iteratively identify notably good and bad features from existing products, and incorporate/remove them from the next generation products. The method proposed in this paper could be incorporated into such a framework to improve the notable product feature extraction process. Sections

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

3.1–3.4will discuss each component in Fig. 1 indetail.

3.1 Data Preprocessing: Social media provides a means for people to interact, share, and exchange information and opinions in virtual communities and networks [42]. Generalization needs minimal assumption about functionalities like a tuple of unstructured textual content, a user ID, and a timestamp which is termed as a message. This minimal assumption would allow the proposed methodology to generalize across multiple heterogeneous pools of social media such as Twitter, Facebook, and Google , as each of these social media platforms has this common data structure. Social media messages, corresponding to each product domain, are retrieved by a query of the product's name (and its variants) within the large stream of social mediadata

Data Cleaning. Most social media crawling application program interfaces provide additional information with each social media message such as user identification,geographical information, and other statistics.Though this additional information could be useful, it is disregarded and removed not only to save storage space and improve computational speed but also to preserve the minimal assumption about the social media data mentionedearlier.

Raw social media need data preprocessing. In order to remove such noise, the data cleaning process does the following:

Lowercasing the textualcontent.

Removing hashtags, usernames, andhyperlinks.

Removing stopwords.

Note that misspelled words (e.g., hahaha and lovin) and emogies are intentionally preserved. Even though they are not well-formed and do not exist in traditional dictionaries, they have been shown to carry useful information that infers semantic meaning behind the messages[8,43]. Since stemming reduce high dimensionality data, it is not utilized in the proposed method.

3.2 Sentiment Extraction. The technique developed by Thelwall et al. is employed to quantify the emotion in a message [43]. The algorithm takes a short text as an input, and outputs two values, each of which ranges from 1 to 5. The first value represents the positive sentiment level, and the other represents thenegative sentiment level. The reason for having the two sentiment scores instead of just one (with –/ sign representing negative/ positive sentiment) is because research findings have determined that positive and negative sentiments can coexist [46]. However, in this research, we only focus on the net sentiment level; hence, the positive and negative scores are combined to produce an emotion strength score.

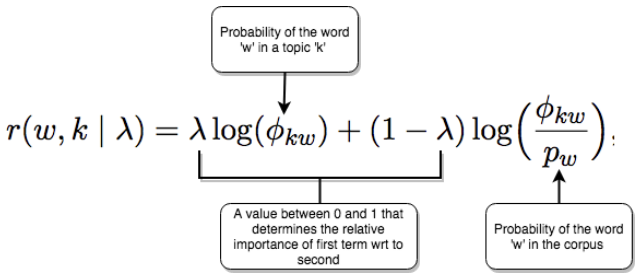
A message is then classified into one of the three categories based on the sign of the emotion strength score (i.e.,positive (1ve), neutral (0ve), and negative (2ve)). The EmotionStrength scores will later be used to identify whether a particular message conveys a positive or negative attitude toward a particular product or product feature in . An example analysis as per statistics on table 4

Table1 – Normal feature extraction without any translation

Method	Baseline								
class	Negative (—)			Neutral (0)			Positive (þ)		
Product	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
HTC	1	0.2	*0.33*	0.25	0.5	0.33	*0.5*	*0.5*	*0.5*
Motorola	0.25	0.5	0.33	0.33	0.2	0.25	0.2	0.27	0.23
Samsung	*0.48*	*0.45*	*0.47*	0.54	0.41	0.47	0.31	*0.5*	*0.38*
iPhone 3	*0.5*	0.42	0.46	0.75	0.4	0.52	0.18	*0.67*	*0.29*
iPhone 4	*0.56*	0.54	0.55	0.48	0.34	0.4	0.59	*0.73*	*0.66*
iPhone 5	0.23	0.26	0.25	0.54	0.33	0.41	0.31	*0.53*	*0.39*
Avg	*0.5*	0.4	0.4	0.48	0.36	0.4	0.35	*0.53*	*0.41*

3.3 Feature Extraction. Product features are extracted from each social media message. In this paper, the feature extraction algorithm used in Ref. [4] is employed. Atahighlevel,thefeature extraction algorithm takes a collection of social messages corresponding to a product as input, and outputs a tuple of rfeature; frequency λ such as h'onscreen keyboard';

The features are extracted because the proposed methodology infers explicit opinions toward a particular product feature; hence it is imperative that product features can automatically be identified



.

3.4 Part of Speech Tagging. The final step of the social media data preprocess is to tag each word in a social message with a part of speech (POS). In this paper, Carnegie Mellon ARK Twitter POS taggeris used for this purpose.

Thepartofspeechinformationisneededinordertodisambiguate words with multiple meanings (i.e., homonyms) [47],which canbecommonlyfoundinsocialmedia.Forexample,theword"super" in mobile context and food context is different.

Each POS tag will become a node type in the coword network. Besides standard linguistic POS tags offered by the POS tagger tool, a special node type PRODUCT is also introduced to distinguish a word that represents a product name from other words.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

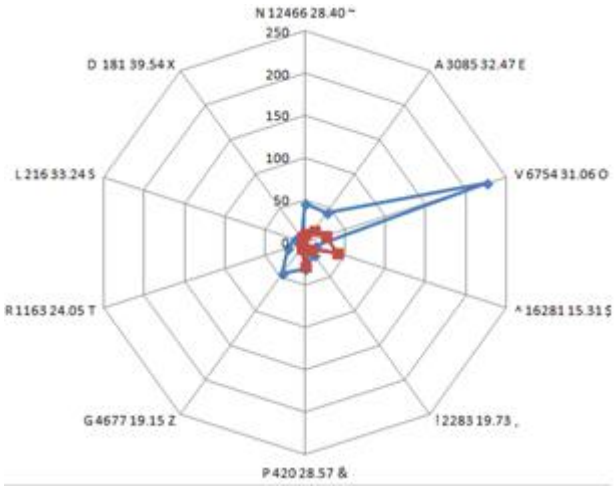
Table 1-Node types and itsdescriptions.

!	Interjection
$	Numeral
&	Coordinating conjunction
,	Punctuation
^	Proper noun
~	Discourse marker, indications of continuation across multiple tweets
A	Adjective
D	Determiner
E	Emotion
G	Other abbreviations, foreign words, possessive endings, symbols, garbage
L	Nominal verbal (e.g., im), verbal nominal (lets)
N	Common noun
O	Pronoun (personal/WH; not possessive)
P	Pre- or postposition, or subordinating conjunction
R	Adverb
T	Verb particle
V	Verb including copula, auxiliaries
Z	Proper noun possessive

3.5 Generating and Indexing Coword Network. A coword network is the collective interconnection of terms based on their paired presence within a specified unit of text. Traditional coword networks represent a node with only textual representation of a word. Variants of co-occurrence networks have been used extensively in the information retrieval field in a wide range of applications that involve semantic analysis such as concept/trend emergence detection [48,49], finding new words, clustering related items [50,51], semantic interpretation [7,52], and document annotation[53,54].
In this paper, a node also incorporates part of speech information for word-sense disambiguation purposes. Concretely, a coword network is an undirected, weighted graph where each node is a pair of hWord+POStag

The coword generation algorithm from a collec- tion of social media messages identifies coherent and incoherent aspects:
For example,
Coherent Aspect:
The martinis were very good.
The drinks both wine and martinis were tasty.
Incoherent Aspect:
Mithra is the best I've ever seen.
It's the best place I've ever seen.

3.6 Sarcasm Detection. A majority of implicit social
Figure 2 – Textual Analysis for Feature Extraction media data are manifested in the form of sarcastic messages. Maynard and Greenwood reported that roughly 22.75% of socialmedia dataissarcastic[14].Hence,thisworkfocusesonimprovingthe ability to interpret sarcastic product-related social media mes- sages. In the proposed framework, sarcastic messages are auto- matically discovered using a machine learning based sarcasm detection algorithm, implemented in Ref. [55]. The algorithm pro- duces a sarcastic message detection model using the features extracted from the training data. These feature sets include the following:



i) n-grams: This feature set extracts individual words (uni-grams) and two consecutive words (bi-grams) from a given message. These n-gram features are used extensively to train classification models for text classification tasks. Three and more consecutive words are not used since research has shown the combination of uni-grams and bi-grams are sufficient and optimal that yields the best results while consuming reasonable amounts of computing resources and memory [56].

ii) Sentiment: It is a hypothesis that sarcastic messages aremore negative than nonsarcastic ones. Moreover, studies show that sarcastic messages tend to exhibitthe co-existence of positive and negative sentiments [46]. The sentiment features include (1) a positive and a negative sentiment score to each word in the message using the SentiWordNet5dictionary, and (2) the sentiment score produced by the python library TextBlob.
iii) Topics: The topical features are extracted using the Latent Dirichlet Allocation algorithm [33] implemented in gensim.The training dataset includes 20,000 sarcastic tweets and 100,000 nonsarcastic tweets. Once the features are extracted from the training data, they are used to train a support vector machine classification model. The trained model is then used to identify a message whether it is sarcastic or nonsarcastic.

3.7 Query and Result Processing. A query is a free text message with implicit content. Inparticular,inorderto processafreetextqueryQText,thefollowingstepsareperformed:
Preprocess the query QText using the mechanism described in Sec. 3.1, in order to clean the raw message, extract features, and assign POStags.

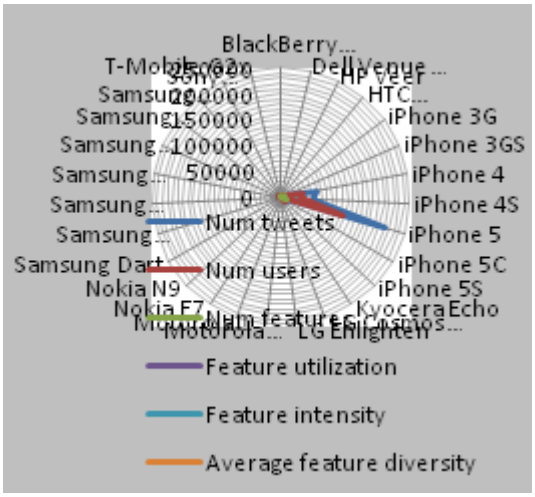
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Form the query compound Q, by converting each POS tagged word into a node, and combining them into a set.
Remove the nodes in Q that do not exist in the coword network.
The resulting query compound Q is then fed into the system for further processing.
The implicit message translation problem in transformed into a node ranking problem so that traditional information retrieval techniques can be applied. In this context, a node in the coword network is equivalent to a combination of a word and its POS.The final output of the system would then be the top words classified by their parts of speech in table 2. Based on the weightage of the expression in sentiment analysis, output classifies the node as highest degree nodes and lowest degree nodes.
For example, a user review can be classified as follows:

IV.CASE STUDY, RESULTS, ANDDISCUSSION

This section introduces a case study used to verify the proposed methodology and discusses the results.
Social Media Data Collection. Twitteris a microblog service that allows its users to send and read text messages of up to 140 characters, known as tweets. The Twitter dataset used in this research was collected randomly using the provided Twitter application program interfaces, and comprises 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year as on Jan 2019 status.

Figure 2 – Tweets processed by smartphone based on Feature Utilization



Most traditional semantic interpretation techniques including sentiment analysis assume that documents are explicit and would fail when dealing with these implicit social media messages. The Column "Sentiment Level (From Implicit Context)" shows quantified sentiment level on the original tweets. The actual Emotional Strength scores are in parentheses. The Column "Manual Sentiment Evaluation" lists the manual evaluation by the authors on the actual sentiment that each
also interesting to note that the sentiment computed for the implicit sample messages tend to be neutral (Sentiment Level 0), regardless of the fact that they are composed with emotion- inspired words (i.e., love, can't, shit, beautifully, and incredible).This agrees with prior findings that messages with implicit sentiment (i.e., sarcasm) would be sentimentally neutralized since suchmessages tend to have equally high volumes of both Positive and Negative scores, causing the Emotion Strength score to converge to 0 [59].
Table 5- sample tweet infers toward the target product features (either Positive or Negative)

Method	Coword								
class	Negative (—)			Neutral (0)			Positive(þ)		
Product	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
HTC ThunderBolt	*0.23*	*0.22*	0.23	*0.52*	*1*	*0.69*	0.3	0.34	0.32
Motorola Droid	*0.46*	*1*	*0.63*	*0.84*	*0.81*	*0.82*	*0.34*	*0.32*	*0.33*
Samsung Galaxy	0.38	0.44	0.41	*0.73*	*0.84*	*0.78*	*0.56*	0.38	*0.45*
iPhone 3	0.41	*0.53*	*0.46*	*0.78*	*0.78*	*0.78*	*0.34*	0.33	*0.34*
iPhone 5	*0.37*	*0.55*	*0.44*	*0.75*	*0.72*	*0.73*	*0.72*	0.37	*0.49*
iPhone 4	0.53	*0.59*	*0.56*	*0.52*	*0.83*	*0.64*	*0.87*	0.38	0.53
Avg	0.4	*0.55*	*0.45*	*0.69*	*0.83*	*0.74*	*0.52*	0.35	*0.41*

. The Column "Sentiment Level (From Translated Explicit Context)" shows the sentiment level using the same sentiment extraction algorithm, but on the translatedexplicit content generated by concatenating the top 20 keywords returned by the system into a single text (disregarding parts of speech). The sentiment levels computed on the translated text agree with the manual evaluation in all the samples shown in Table 7.
Not surprisingly, the sentiment level extracted from the original also interesting to note that the sentiment computed for the implicit sample messages tend to be neutral (Sentiment Level 0), regardless of the fact that they are composed with emotion- inspired words (i.e., love, can't, shit, beautifully, and incredible).This agrees with prior findings that messages with implicit senti- ment (i.e., sarcasm) would be sentimentally neutralized since suchmessages tend to have equally high volumes of both Positive and Negative scores, causing the Emotion Strength score to convergeto 0 [59]. If text is all incorrect, since the sentiment extraction technique is designed to detect explicit sentiment, and hence would not give correct results when dealing with sarcasm or vague context,
Similarly the coword network relatively improves the estimation of textual analysis as shown in fig. 3 where p denotes precision, R represents Recall and F represents F-measure.
Figure 3- Comparison of baseline extraction of positive values (without any translation)

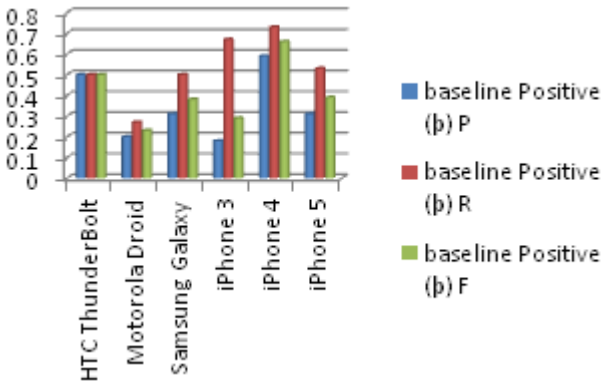


Figure 4- Performance of coword networks in extracting positive context

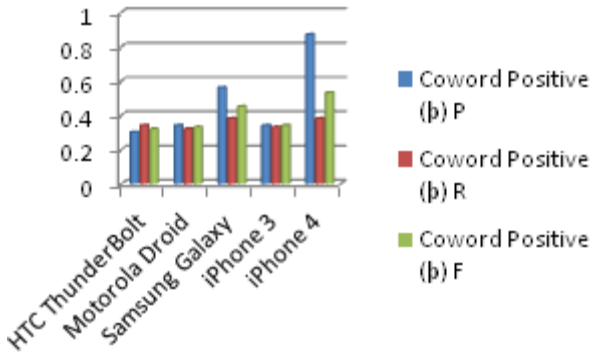
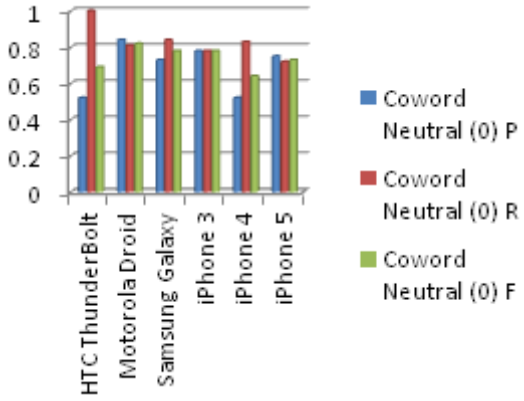


Figure5- Improvements in evaluation of neutral values



III. CONCLUSIONS AND FUTUREWORKS

This paper proposes a knowledge-based methodology for inferring explicit sense from social media messages whose connotations related to products/product features are implicit. The methodology first generates a coword network from the corpus of social media messages, which is used as the knowledge source that captures the relationship among all the words present in the social To know the customers' view, an objective type of questionnaire was prepared and distributed to them who had minimum two online purchasing experiences. The questionnaire was handed over and collected the filled questionnaire personally. The objective of this study is to find out the young customers buying behaviour and experience regarding their day to day online shopping experiencew.r.t different aspects of online marketing. The questionnaire provides the customer an opportunity to express their views and concerns which they face on a regular basis while buying through online. This study will help the marketers to identify the challenges affecting buying behaviour of online customers and to identify the areas where these marketers need to formulate the future policy that further helps in retentionof young customers. The survey reveals a number of interesting facts when we interviewed the respondents. Selected customers in tier II cities namely Madurai, Tiruchirappalli and Coimbatore in Tamil Nadu were the respondents. Let us analyze few responses, which can be taken as a strong indicator for awareness of online marketing, its popularity over time and young customers'involvement towards it.

REFERENCES

1. Tuarob, S, Tucker, C. S. 2015, "Automated Discovery of Lead Users and Latent Product Features by Mining Large Scale Social Media Networks" ASME J. Mech. Des., 137(7), p. 071402.
2. Tuarob, S., and Tucker, C. S., 2015, "Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data," ASME J. Comput. Inf. Sci. Eng., 15(3), p. 031003.1485–1509.
3. Tuarob, S., and Tucker, C. S., 2015, "A Product Feature Inference Model for Mining Implicit Customer Preferences Within Large Scale Social Media Networks," ASME Paper No. DETC2015-47225.
4. Tuarob, S., and Tucker, C. S., 2014, "Discovering Next Generation Product Innovations by Identifying Lead User Preferences Expressed Through Large Scale Social Media Data," ASME Paper No. DETC2014-34767.
5. Tuarob, S., and Tucker, C. S., 2013, "Fad or Here to Stay: Predicting Product Market Adoption and Longevity Using Large Scale, Social Media Data," ASME Paper No. DETC2013-12661.
6. Lim, S., and Tucker, C. S., 2016, "A Bayesian Sampling Method for Product Feature Extraction From Large-Scale Textual Data," ASME J. Mech. Des., 138(6), p. 061403.
7. Tuarob, S., Tucker, C. S., Salathe, M., and Ram, N., 2014, "An Ensemble Het- erogeneous Classification Methodology for Discovering Health-Related Knowl- edge in Social Media Messages," J. Biomed. Inf., 49, pp. 255–268.
8. Tuarob, S., Tucker, C. S., Salathe, M., and Ram, N., 2013, "Discovering Health-Related Knowledge in Social Media Using Ensembles of Heterogeneous Features," 22nd ACM International Conference on Information & Knowledge Management (CIKM '13), San Francisco, CA, Oct. 27–Nov. 1, pp. 1685–1690.
9. Lim, S., Tucker, C. S., and Kumara, S., 2017, "An Unsupervised Machine Learning Model for Discovering Latent Infectious Diseases Using Social Media Data," J. Biomed. Inf., 66, pp. 82–94.
10. Sakaki,T.,Okazaki,M.,andMatsuo,Y.,2010,"EarthquakeShakesTwitter Users: Real-Time Event Detection by Social Sensors," 19th International Con- ference on World Wide Web (WWW'10), Raleigh, NC, Apr. 26–30, pp. 851–860.
11. Caragea,C.,McNeese,N.,Jaiswal,A.,Traylor,G.,Kim,H.,Mitra,P.,Wu,D., Tapia, A., Giles, L., Jansen, B., and Yen, J., 2011, "Classifying Text Messages for the Haiti Earthquake," Eighth International Conference on Information Sys- tems for Crisis Response and Management (ISCRAM), Lisbon, Portugal, May 8–11.
12. Bollen, J., Mao, H., and Zeng, X., 2011, "Twitter Mood Predict
13. the Stock Market," J. Comput. Sci., 2(1), pp.1–8.
14. Zhang, X., Fuehres, H., and Gloor, P., 2012, "Predicting Asset Value Through Twitter Buzz," Advances in Collective Intelligence 2011, Springer, Berlin, pp. 23–34.
15. Maynard, D., and Greenwood, M. A., 2014, "Who Cares About Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis," Ninth International Conference on Language Resources and Evaluation (LREC), Rey- kjavik, Iceland, May 26–31, pp.4238–4243.
16. Dey, L., and Haque, S., 2009, "Studying the Effects of Noisy Text on Text Min- ing Applications," Third Workshop on Analytics for Noisy Unstructured Text Data (AND), Barcelona, Spain, July 23–24, pp. 107–114.
17. Tsur, O., Davidov, D., and Rappoport, A., 2010, "ICWSM-A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews," Fourth International Conference on Weblogs and Social Media (ICWSM), Washington, DC, May 23–26, pp.162–169.
18. Davidov, D., Tsur, O., and Rappoport, A., 2010, "Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon," 14th Conference on Computa- tional Natural Language Learning (CoNLL), Uppsala, Sweden, July 15–16, pp. 107–116.
19. Navigli, R., and Velardi, P., 2005, "Structural Semantic Interconnections: A Knowledge-Based Approach to Word Sense Disambiguation,"

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IEEE Trans.Pat-Gen., 115(1), p. 3.
20. Gibbs, R. W., and Colston, H. L., 2007, Irony in Language and Thought: A Cognitive Science Reader, Lawrence Erlbaum, NewYork.
21. Archak, N., Ghose, A., and Ipeirotis, P. G., 2011, "Deriving the Pricing Power ofProductFeaturesbyMiningConsumerReviews,"Manage.Sci.,57(8),pp.
22. Asur, S., and Huberman, B. A., 2010, "Predicting the Future With Social Media," IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Washington, DC, Aug. 31–Sept. 3, pp. 492–499.
23. Stone, T., and Choi, S.-K., 2014, "Visualization Tool for Interpreting User Needs From User-Generated Content Via Text Mining and Classification," ASME Paper No. DETC2014-34424.
24. Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., and Li, X., 2011, "Comparing Twitter and Traditional Media Using Topic Models," Advances in Information Retrieval, Springer, Berlin, pp. 338–349.
25. Yajuan, D., Zhimin, C., Furu, W., Ming, Z., and Shum, H. Y., 2012, "Twitter Topic Summarization by Ranking Tweets Using Social Influence and Content Quality," 24th International Conference on Computational Linguistics, Mumbai, India, Dec. 8–15, pp. 763–780.
26. Wang, Y., Wu, H., and Fang, H., 2014, "An Exploration of Tie-Breaking for Microblog Retrieval," Advances in Information Retrieval, Springer, Cham, Switzerland, pp. 713–719.
27. Tuarob, S., Tucker, C. S., Salathe, M., and Ram, N., 2015, "Modeling Individual-Level Infection Dynamics Using Social Network Information," 24th ACM International on Conference on Information and Knowledge Manage- ment, Melbourne, Australia, Oct. 19–23, pp. 1501–1510.
28. Tuarob, S., and Mitrpanont, J. L., 2017, "Automatic Discovery of Abusive Thai Language Usages in Social Networks," International Conference on Asian Digi- tal Libraries, Bangkok, Thailand, Nov. 13–15, pp. 267–278.
29. Thelwall, M., Buckley, K., and Paltoglou, G., 2011, "Sentiment in Twitter Events," J. Am. Soc. Inf. Sci. Technol., 62(2), pp. 406–418.
30. Thelwall, M., 2017, "The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web With SentiStrength," Cyberemotions, Springer, Cham, Switzerland, pp. 119–134.
31. Tuarob, S., Tucker, C. S., Kumara, S., Giles, C. L., Pincus, A. L., Conroy, D. E., and Ram, N., 2017, "How are You Feeling?: A Personalized Methodology for Predicting Mental States From Temporally Observable Physical and Behav- ioral Information," J. Biomed. Inf., 68, pp.1–19.
32. Fox, E., 2008, Emotion Science: Cognitive and Neuroscientific Approaches to Understanding Human Emotions, Palgrave Macmillan, Basingstoke,UK.
33. Cutting, D., Kupiec, J., Pedersen, J., and Sibun, P., 1992, "A Practical Part-of- Speech Tagger," Third Conference on Applied Natural Language Processing (ANLC '92), Trento, Italy, Mar. 31–Apr. 3, pp.133–140.
34. 33.33.333333.Jia, S., Yang, C., Liu, J., and Zhang, Z., 2012, "An Improved Information Fil- tering Technology," Future Computing, Communication, Control and Manage- ment, Springer, Berlin, pp.507–512.
35. Tuarob,S.,Mitra,P.,andGiles,C.L.,2012,"ImprovingAlgorithmSearchUsing the Algorithm Co-Citation Network," 12th ACM/IEEE-CS Joint Conference on DigitalLibraries(JCDL'12),Washington,DC,June10–14,pp.277–280.
36. Tuarob, S., Bhatia, S., Mitra, P., and Giles, C., 2013, "Automatic Detection of Pseudocodes in Scholarly Documents Using Machine Learning," 12th Interna- tional Conference on Document Analysis and Recognition (ICDAR), Washing- ton, DC, Aug. 25–28, pp. 738–742.
37. Evans, D. A., Handerson, S. K., Monarch, I. A., Pereiro, J., Delon, L., and Hersh, W. R., 1998, Mapping Vocabularies Using Latent Semantics, Springer, Boston, MA.
38. Tuarob, S., Pouchard, L. C., and Giles, C. L., 2013, "Automatic Tag Recom- mendation for Metadata Annotation Using Probabilistic Topic Modeling," 13th ACM/IEEE-CS Joint Conference on Digital Libraries, (JCDL'13), Indianapolis, IN, July 22–26, pp.239–248.
39. Tuarob, S., Pouchard, L., Mitra, P., and Giles, C., 2015, "A Generalized Topic Modeling Approach for Automatic Document Annotation," Int. J. Digital Libr., 16(2), pp. 111–128.
40. Cliche,M.,2014,"TheSarcasmDetector:LearningSarcasmFromTweets!," The SarcasmDetector,accessedFeb.19,2017,http://www.thesarcasmdetector.com
41. Liu,F.,Liu,F.,andLiu,Y.,2008,"AutomaticKeywordExtractionfortheMeet- ing Corpus Using Supervised Approach and Bigram Expansion," Spoken LanguageTechnologyWorkshop(SLT2008),Goa,India,Dec.15–19,pp.181–184.
42. Martin,S.,Brown,W.M.,Klavans,R.,andBoyack,K.W.,2011,"OpenOrd: An Open-SourceToolboxforLargeGraphLayout,"SPIEProc.,7868,p.786806.
43. Tuarob, S., Pouchard, L. C., Noy, N., Horsburgh, J. S., and Palanisamy, G., 2012, "Onemercury: Towards Automatic Annotation of Environmental Science Metadata," Second International Workshop on Linked Science, Boston, MA, Nov.12.