Procedia Computer Science

Volume 53, 2015, Pages 199–207

2015 INNS Conference on Big Data

2015 INNS Conference on Big Data

Ontology-based Sentiment Analysis Process for Social Media Content

Pratik Thakor[a], Dr. Sreela Sasi[b]

[a]*1605 Arizona Ave, Milpitas, 95035, USA*
[b]*109 University Square, Erie, 16501, USA*

Abstract

Social media provides a platform where users share an abundance of information on anything and everything. The information may consist of users' emotions, feedbacks, reviews, and personal experiences. In this research a novel Ontology-based Sentiment Analysis Process for Social Media content (OSAPS) with negative sentiments is presented. The social media content is automatically extracted from the twitter messages. An ontology-based process is designed to retrieve and analyse the customers' tweet with negative sentiments. This idea is demonstrated with the identification of customer dissatisfaction of the delivery service issues of the United States Postal Service, Royal Mail of United Kingdom, and Canada post. The tweets related to the delivery service include delay in delivery, lost package/s or improper customer services at the office in person or at call centres. A combination of technologies for twitter extraction, data cleaning, subjective analysis, ontology model building, and sentiment analysis are used. The results from this analysis could be used by the company to take corrective measures for the problems as well as to generate an automated online reply for the issues. A rule-based classifier could be used for generating the automated online replies.

Keywords: Ontology-based Process, Negative Sentiment Analysis, Subjective Analysis, Web Mining, Text Parsing

doi:10.1016/j.procs.2015.07.295

1. Introduction

The sentiment analysis of customer's social media data is very important in the present day business scenarios. Customers share information about products, services and their experiences on social media. This information can be used for market research, product feedback and analysing customer service effectiveness. Bing Liu explained that opinions are subjective expressions. Opinions describe people's sentiments or feeling towards entities, events or their properties [1]. The sentiment analysis of the opinions could lead to many interesting results. The dynamically expanding web and social media are generating huge amount of opinion data. People's opinion about any product or service on social media is a very valuable asset for any organization. The organizations can generate information on customer's response, or its behaviour for any product or service, by doing the sentiment analysis of these social media data.

Mike Thelwall, Kevan Buckley, Georgios Paltoglou, and Di Cai presented an algorithm for understanding the sentiment associated with the short text [2]. The users' shared informal text or sentences convey different emotions on 'likes' and 'dislikes'. They developed algorithms to identify the sentiment and the sentiment strength in these informal communications. Janyce Wiebe and Ellen Riloff showed a technique for creating subjective and objective sentence classifiers from unannotated texts [3]. They tried to develop a system to analyse text and find out opinion or emotions associated with it. They suggested a method for classifying subjective and objective information from a sentence. This subjective and objective information can be used to develop different models, which can be utilized to answer a customer based on their behaviour.

In this research, a process for sentiment analysis of customers' social media data using an ontology model is presented. This process would help in identifying the problem area associated with the customers' social media data that contains negative sentiments. The remaining portion of the paper is organized as follows: Section II provides the background research used for building the ontology model, Section III provides information on building an ontology model and the use of ontology model to for sentiment analysis, Section IV focuses on discussion about different attempts taken during this research, their limitation and usefulness, Section V provides the Simulation Results and Section VI briefs on conclusions and future work.

2. Background Research

"Ontology" represents the domain knowledge as a hierarchy of concepts. It includes machine-interpretable definitions of basic concepts of the domain and the interrelation of these concepts. Entities (classes), objects, object properties and the relation between them could be incorporated in to an ontology model. Also, it defines a common vocabulary for researchers who need to share information in a domain. Defining ontology helps in sharing a common understanding of knowledge among people and software agents, and makes domain knowledge reusable. The ontology model could be stored in OWL/XML or

RDF/XML format. It is possible to merge two ontology models with the same knowledge domain. We can retrieve the information organized in ontology model by querying it.

Sayed Zeesan Haider and Yaakub, R. M, Li. & Feng, Y. developed an ontology to do feature based sentiment analysis of customer's review on smart phones, and simple mobile phones [4, 5]. K.M Sam and C. R. Chatwin proposed a relatively similar model for doing sentiment analysis of customer reviews on electronic products [6]. Larissa A. de Freitas and Renata Vieira had also built and used an ontology model to do feature based opinion mining of Portuguese movie reviews [7]. All these research works used the information on categories and features of the products to build the ontology model. Later the ontology model was used to identify the most popular products by analyzing the customer reviews. Tim Finin, Li Ding and Lina Zou shared an interesting concept of developing an ontology based intelligent application [8]. Matteo Baldoni, Cristina Baroglio, Viviana Patti and Paolo Rena presented an ontology based approach for sentiment analysis of social semantic web [9]. The idea was to share information and knowledge in an effective way with the help of ontology. They showed the use of ontologies to enable computer programs to read, to publish or to exchange information and knowledge by enhancing interoperability. Natalya F. Noy and Deborah L. McGuinness from Stanford University gave a perfect insight into the process of building ontology from the scratch [10]. They provided excellent information about the various components of the ontology model. They explained the idea of identifying the important building components of an ontology model in a text by applying OOP concepts.

In this research an ontology model based process for identifying customer dissatisfaction with any postal service from their tweets is developed. The subjective analysis of tweet text is done to retrieve the information on sentence builders like nouns and verbs. By using the OOP concepts, the noun and verb can be used as object and object property respectively. This information on objects and object properties were used to build the ontology model. Mike Thelwall, Kevan Buckley, Georgios Paltoglou, and Di Cai used SentiStrength tool to identify sentiments associated with random text extracted from social media [2]. The SentiStrength can calculate the strength of the sentiments associated with the text. Only negative tweets are analyzed in this research to find the problem area associated with it.

3. Methodology

The methodology consists of two processes. The first process is to build the ontology model using the data extracted from the social media platform, Twitter. The second process is to retrieve the problem area from the negative sentiments associated with a tweet using a previously built ontology model. These processes have tasks such as social media data extraction and data cleaning, identifying negative sentiment in tweets text, subjective analysis, building ontology model, query building, and retrieving information from the ontology model.

3.1 Building an Ontology model

The process of building an ontology model is shown in Fig. 1. The data is to be extracted from the social media platform, Twitter. The script written in python is used to extract the tweets for USPS, Royal mail and Canada post and to clean it. Again data cleaning is performed with macros on the tweets to remove tweets with hyperlinks or any other special characters. The postal service domain with its class, object and object properties are used to build ontology model. H. Cunningham, et al. showed a methodology to do text parsing using GATE software [11]. The subjective analysis of tweet data was done to identify the objects and their object properties for the postal service domain. A combination of NLP-based language parsing plugins in GATE was used for annotating the nouns and verbs in the tweet. The output of the GATE software was in a tag form with nouns and verbs enclosed in tags. Data cleaning was done to get the nouns and verbs from the results. The script written in Python was used for data cleaning. The data after cleaning had redundancies of noun and verb. These redundancies were removed by using Excel Macros. The final results had only nouns and verbs from tweet texts.

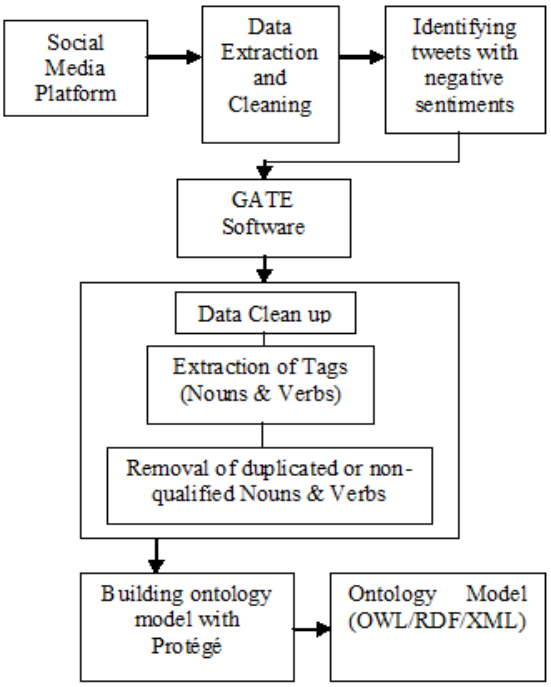

Fig. 1: Ontology model building process

The Object Oriented Programming (OOP) concepts were applied on this result. The nouns are considered as objects and the verbs as object properties. These classes, objects and object properties' information are used to build the ontology model. The Stanford's

Protégé software is used to build the ontology model. Class, object and object property are identified as entity, individual, and object property in the ontology model respectively. The relations between classes, objects and object properties were derived manually as per the human understanding of a sentence. For E.g. In postal service domain, it is understandable that a mail or letter could be delivered. So "mail and letter" are identified as objects and are assigned to the class called "Delivery". Generally mail or letter is handled by postman. "Postman" can be assigned to the "Field worker" class. In lost mail, wrong delivery of mail, or delayed delivery of mail, the "Lost", "Wrong delivery", and "Delayed delivery" are used as object property for the object "Mail" and the class "Delivery". The postman as a "Field worker" is also associated with them. In some scenarios a package could be damaged while handling or delivering. In such case the "package", an object, is assigned to the class 'Package' and object property of "handling and damaging" is assigned to the object "package".

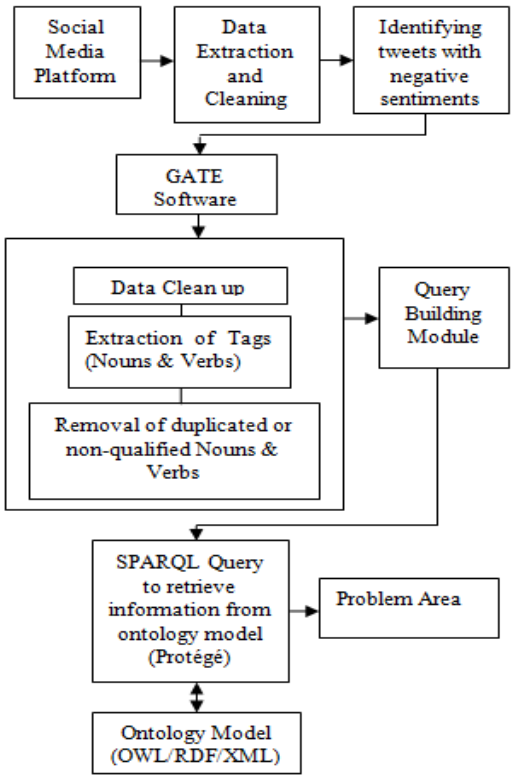

Fig. 2: Process of retrieving information from ontology model

3.2 Sentiment analysis using the ontology model

The information retrieval process from the ontology model is shown in Fig. 2. The Protégé software is used to retrieve information from the ontology model. The previously built ontology model on customer's postal services related issues is used for querying and identifying the problem area associated with customer's tweet. The tweets with the negative sentiments are used to do sentiment analysis and to find the problem area associated with the customer issues. SentiStrength tool was used to identify the tweet with the negative sentiments. Fig. 3 shows the results from the SentiStrength tool. The tool was used for identifying the polarity of words in the tweet sentences. The negative polarity corresponds with negative sentiments and positive polarity corresponds with positive sentiment. The tweets with their higher negative polarity compared to their positive polarity were considered as tweets with negative sentiments.

The important point is, if a tweet has two sentences in it then the noun and verb results for each sentence should be handled separately. It is logical that two different sentences should represent two different perspectives sometimes. Only a combination of valid noun and verb is considered to build a query. Only noun or only verb is not considered for building the query. Fig. 4 shows a tweet and its GATE results for noun and verbs. The GATE results can have some irrelevant nouns and verbs. E.g., the results like 'one, 'we', 'they' 'did' and 'it'. These results were not considered for building the query or to retrieve information from the ontology. Fig. 5 shows an iteration for identifying the query data from a tweet. This association approach for identifying the query data is not an efficient one. It is suggested to consider other irrelevant information that is omitted here, and to build more perfect sentence based logic to build the SPARQL query.

```
USPS didn't deliver one of the Xmas packages we had sent here.
Trying to figure out wtf they did with it #HateUSPS
1   -3   USPS[0] didn't[0] deliver[0] one[0] of[0] the[0] Xmas[0]
packages[0] we[0] had[0] sent[0] here[0]
[[Sentence=-1,1=word max, 1-5]] Trying[0] to[0] figure[0] out[0]
wtf[-2] they[0] did[0] with[0] it[0] #HateUSPS[0]
[[Sentence=-3,1=word max, 1-5]][[[1,-3 max of sentences]]]

Slept through my Amazon USPS delivery :( I HATE that! Maybe that
guy is right and I SHOULD slit my wrists!   Slept through my Amazon
USPS delivery delivery I HATE that!<br> Maybe that guy is right and
I SHOULD slit my wrists!
1   -4   Slept[0] through[0] my[0] Amazon[0] USPS[0] delivery[0]
[-1 Emoticon] I[0] HATE[-3] that[0] [[Sentence=-4,1=word max, 1-5]]
Maybe[0] that[0] guy[0] is[0] right[0] and[0] I[0] SHOULD[0] slit[0]
wrists[0] [[Sentence=-1,1=word max, 1-5]][[[1,-4 max of sentences]]]
```

Fig. 3: SentiStrength result of negative sentiment analysis for sample tweets

USPS didn't deliver one of the Xmas packages we had sent here.
Trying to figure out wtf they did with it #HateUSPS

```
<paragraph>
<NP>USPS</NP> <VP>didn&apos;t deliver</VP> <NP>one</NP> of
<NP>the Xmas packages</NP> <NP>we</NP> <VP>had sent</VP> here.
<VP>Trying to figure</VP> out wtf <NP>they</NP> <VP>did</VP> with
<NP>it</NP> #HateUSPS
</paragraph>
```

Fig. 4: GATE result for a sample tweet

Iteration 1			**Iteration 2**	
NP	**VP**		**NP**	**VP**
USPS	didn't deliver		the Xmas packages	didn't deliver
the Xmas packages	had sent			had sent
one				
we				Trying to figure
				did
they	Trying to figure			
it	did			
Iteration 3			**Final Results**	
NP	**VP**		**Combination**	
the Xmas packages	didn't deliver		the Xmas package, didn't deliver	
	had sent		the Xmas package, had sent	
	Trying to figure			
	did			

Fig. 5: An iteration for identifying the query data from a tweet

5. Results

The result obtained using the Ontology based Sentiment analysis is shown in Table 1.The clean-up process removed the hyperlinks, image, and any non-English tweet, from the final data set. Only the tweets with negative sentiments were used for this experiment. The intention was to find out reasons for customers dissatisfaction. The ontology model that was built earlier was used to identify the services/products for which customers had published negative tweets. The Ontology model was queried using SPARQL with the input as different combinations of nouns and verbs extracted from these negative tweets.

The results can be improved by refining relations between objects in ontology model. The ontology model used for this experiment had less than 100 objects and was built with information from around 250 tweets.

Table 1. Results of Successful Analysis

	USPS	**Royal mail**	**Canada post**
Tweets used for Experiment	250	250	250
Tweets used after clean up	158	182	154
Tweets with negative sentiments	78	121	54
Tweets with positive sentiments	31	49	87
Tweets with neutral sentiments	49	12	13
Success in identifying product/service in tweet	22	21	36

6. Conclusion and Future Work

This research focused on building an ontology model to identify the divisions or sections of the office having problem areas associated with customers' dissatisfaction in the Postal Service. This is done by analyzing their social media content. A partially automatic process was developed to find out the problem area associated with the content shared on social media with the help of an ontology model. Though the modules worked as expected, it required considerable amount of manual effort in terms of identification of relations between objects and properties. Data entry in Protégé software to build an ontology model, and data cleaning modules were executed manually. Few modules such as subjective analysis with GATE software and SPARQL query building need further work in order to optimize the process and to get accurate results. The system sometimes gives results that don't have any logical association with the content in the current process. The results of the GATE software as well as non-detailed ontology model are responsible for it. The ontology model needs to be refined as much as possible. It needs to keep updated with the new information about classes, objects and object properties. The assigning of 'objects' to the 'class' and object 'property' to the 'object' must be refined. The uniqueness of relation between a class, an object and an object property needs to be maintained while refining the ontology model. It needs more work to automate the process of assigning the object 'properties' with the 'object'. The identification of the tweets with the negative sentiments needs to be done in a more optimized and efficient way. The SentiStrength tool couldn't give the accurate results for sarcastic and philosophical text. The SentiStrength's libraries of words with the polarization scores need to be updated with more new words and their polarization scores. The process of building SPARQL query could be optimized. The current many-to-many association rule to create conditional block of query with noun and verb is not efficient. It ended up using all the different association to build SPARQL

query. It should use minimum number of noun and verb associations. The selection of noun and verb must be done in a way that the use of them in SPARQL query gives only the most accurate problem area in result. The accuracy of result also depends on the uniqueness of interrelations and the knowledge built in the ontology model.

Acknowledgements

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed. We are thankful to GATE developer community, Opinionfinder development team, SentiStrength Development team, and Stanford University for Protégé software.

References

[1] Bing Liu, "Sentiment Analysis and Subjectivity", from Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010.

[2] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, and Di Cai, "Sentiment Strength Detection in Short Informal Text", Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK, 2010.

[3] Janyace Wiebe and Ellen Riloff (2005). "Creating subjective and objective sentence classifiers from unannotated texts." Sixth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2005).

[4] Sayed Zeesan Haider, "Ontology-based sentiment analysis case study", a case study for Master degree project, University of Skovde, pp. 05-67, 2012.

[5] Yaakub, R. M, Li. & Feng, Y., "Integration of Opinion into Customer Analysis Model", proceedings of Eighth IEEE International Conference on e-Business Engineering, pp. 90-95, 2011.

[6] K.M Sam and C.R. Chatwin, "Ontology-Based Sentiment Analysis Model of Customer reviews for Electronic Products", Proceedings of International Journal of e-Business, e-Management and e-Learning, Vol. 3, No. 6, Dec.2013.

[7] Larissa A. de Freitas and Renata Vieira, "Ontology-based Feature Level Opinion Mining for Portuguese Reviews", PUCRS FACIN, Porto Alegre, Brazil, 2013.

[8] Tim Finin, Li Ding and Lina Zou "Social Networking on the Semantic Web", Learning Organization Journal, special issue on Ubiquitous Business Intelligence, Miltiadis Lytras et al, 2005.

[9] Matteo Baldoni, Cristina Baroglio, Viviana Patti and Paolo Rena, "From Tags to Emotions: Ontology-driven Sentiment Analysis in the Social Semantic Web", Universit`a degli Studi di Torino, 2010.

[10] Natalya F. Noy and Deborah L. McGuinness "A guide to creating your first ontology", Stanford University.
http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

[11] H. Cunningham, et al. "Text Processing with GATE (Version 6)". University of Sheffield Department of Computer Science. Apr.2011. ISBN 0956599311.
https://gate.ac.uk/