Ontology-based Sentiment Analysis Model for Recommendation Systems

Samreen Zehra, Shaukat Wasi, Imran Jami, Aisha Nazir, Ambreen Khan and Nusrat Waheed
Department of Computer Science, Mohammad Ali Jinnah University, Karachi, Pakistan

Keywords: Ontology, Sentiment Analysis, Recommendation System.

Abstract: In this paper, we propose a novel approach towards developing a recommendation system using ontology-based sentiment analysis. To conduct our study, we have targeted a Facebook closed group which contains posts/reviews regarding different schools. For elucidating the knowledge domain, a school ontology is manually designed based on a set of extracted post/comment data. Sentiment analysis is consequently performed on the resulting Data set and the relative sentiment scores are stored back in the ontology for making recommendations in future.

1 INTRODUCTION

Sentiments are very important as they make a foundation for decision making for the stakeholders. Opinion mining (also termed as Sentiment analysis) is a sub-field of web mining that deals with extracting, analyzing and classifying the sentiments of different people regarding a target entity and its related attributes, from user-generated reviews(Liu, 2012).

Opinions are studied at three levels of granularity (document, sentence or feature) (Freitas and Vieira, 2013). In our case, sentence level analysis is being used to predict user's sentiment correctly in each comment/post, that has no underlying semantic context. It is important to understand that sentiment analysis can only be performed on subjective sentences that contain some personal sentiment, stance, view, or belief attributed towards an entity. The task of classifying a given text being objective or subjective is referred to as Subjectivity classification(Liu, 2012). In our research, we have selected subjective comments manually from Facebook comments file. SentiStrength tool was then employed for discovering the polarity of a given subjective comment, i.e. neutral, positive or negative and to assign a numeric score that expresses the strength/intensity of the sentiment. Considering the various hurdles in conducting sentiment analysis of unstructured texts, the researchers have started exploring semantic-based approaches for opinion mining. This is mostly achieved by utilizing an ontology. Ontology is a formal explicit description of concepts in a domain of discourse in the form of taxonomic hierarchy (Noy et al., 2001). It develops a shared understanding of domain by building a common vocabulary lexicon. An Ontology comprises of concepts, instances, relationships between the concepts, facets and slots. For implementation of our ontology, we have used Protégé, a JAVA based knowledge base framework that complies with W3C standards of representing ontology in the semantic web.

For opinion mining, it is very challenging to identify the target of the opinion in a given text. This problem is known as Named Entity Recognition in Natural Language Processing. The task of finding the target/feature may be eliminated if the knowledge domain is modeled as an ontology, based on predetermined relationships between domain objects and its features, having specific rules and axioms. During the past few years, the social media has emerged as a pivotal source for opinion mining as different users share their experiences and feedback on it. It is worth noting that mining user's perspective, from these short texts/comments, regarding a particular target is a demanding job given the fact that these comments/reviews are characterized as ambiguous, sarcastic and informal. Presently, so much work has already been done on sentiment analysis of a normal unstructured text corpus using lexicon-based approach and machine learning approach or the combination of two.

In our proposed work, we are trying to study sentiments of a particular Facebook group users towards different schools to formulate a recommendation system based on their reviews. A school ontology has been developed to model the school knowledge-base for extraction of relevant opinions and polarity cal-

Zehra S., Wasi S., Jami I., Nazir A., Khan A. and Waheed N.
Ontology-based Sentiment Analysis Model for Recommendation Systems.
DOI: 10.5220/0006491101550160
In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD 2017)*, pages 155-160
ISBN: 978-989-758-272-1

culation. Hence, it is highly likely that a person will have a recommendation based on its requirement. This paper is organized as follows: Section II presents Related Work; Section III describes The Proposed Approach; Section IV explains the Methodology; Section V highlights Experimental Results and Section V outlines Conclusions and Future Work.

2 RELATED WORK

In the literature, similar work in this domain is carried out by (Kontopoulos et al., 2013) where the researchers applied ontology-based sentiment analysis methodology on twitter data. They worked on semi-automatic ontology generation based on twitter posts about different mobile phones and their associated features. Separate Datasets comprising of Users' reviews were used for ontology creation and sentiment analysis. In their work the sentiment scores were computed using a webservice named OpenDover. However, the idea of application of sentiment analysis on an ontology-based knowledge domain was originally proposed by (Yaakub et al., 2012) which was further analyzed and extended by (Haider, 2012) to include feature-based sentiment analysis of customers' reviews on smartphones having enhanced product feature set. Both of them classified the sentiment polarity on a 7-point scale ranging from 3 to -3. They also asserted that the feature level sentiment scores help in the calculation of the overall sentiment score for the object. In a similar context, feature level opinion mining is performed by (Freitas and Vieira, 2013) for Portuguese movie reviews, by making use of already available movie ontology. Another noticeable work by (Sam and Chatwin, 2015) suggested that the polarity of the sentiment word heavily depends on the context in which it is being used. They, therefore, developed two different ontologies. One is emotion ontology for storing emotion words of customers towards different electronic products with associated polarities, the other is the ontology for electronic product domain. (Thakor and Sasi, 2015) applied sentiment analysis on negative tweets for a postal service to figuring out the reason for customer dissatisfaction. In his research, he practiced sentimental analysis using SentiStrength tool and data was retrieved using Protégé SPARQL from ontology.(Alkadri and ElKorany, 2016) used NLP techniques like tokenization, De-noise, Stemming and POS-Tagging for data preprocessing and combined three (3) different Arabic polarity lexicons to form a large-scale Arabic opinion lexicon (ArOpL) to identify Polarity in Arabic language. As mentioned by (Binali et al., 2009), opinion mining research comprises of both feature-level opinion mining and its related sentiment classification. Therefore, two similar objects can be compared based on their overall sentiment scores. Likewise, feature level comparison may also be made between two objects based on their corresponding feature level sentiment scores.

In the current state-of-the-art, ontology is being employed for sentiment analysis but the work is not being extended to frame a recommendation system. Hence we claim in this research that our effort is different from similar work already done in this domain as we store the calculated sentiment scores in the same ontology on corresponding hierarchical levels(i.e. class level and object level) for making recommendations. In this way, we can avoid recalculation of sentiment scores each time the user submits a query. However, it requires ontology modification after its formal conceptualization. It is assumed that the proposed system may also be applied to different domains after careful modifications in ontology and sentiment lexicon. Table 1 presents a detailed comparison of the related work already done in the domain with our work.

3 PROPOSED APPROACH

In our proposed approach the ontology is engineered using user comments. We have deliberately selected a specific group where people share their feedback regarding different schools. It helped us in simplifying the task of subjectivity classification. After extracting reviews regarding target school branches in a file, sentiment analysis is performed and the resulting sentiment scores are stored back into the ontology. Later on, the user can seek a recommendation from the system through a user interface where he/she can find the best school based on his/her requirements(like location, gender-wise orientation).The back-end of the interface translates user request into a query to get the branch level/school level scores from the ontology and gives a recommendation after comparison of sentiment scores.

Figure 1 depicts the proposed approach to our research problem having the sequence below.

1. A school domain ontology is manually created after reviewing Facebook posts/comments using Protégé modeling environment.
2. Using ontology created in previous step, respective objects/classes are extracted which are targets for our sentiment analysis.
3. Opinionated comments/posts are extracted from

Table 1: GAP Analysis.

Research paper	Ontology creation	Sentiment classification Approach	Querying Ontology	Making Recommendation
(Kontopoulos et al., 2013)	An ontology engineer developed ontology manually based on sample tweets extracted from twitter. Another alternative semi-automatic approach is also discussed using OntoGen tool	OpenDover web service is used for calculating polarity score	Not discussed	Not discussed
(Thakor and Sasi, 2015)	Data is extracted from twitter and cleansed for input into GATE software. Results of GATE software are used for manual Ontology creation in Protege	Only negative tweets are considered for Sentiment analysis. Sentiment polarity is assigned using SentiStrength tool.	SPARQL for Protégé is used for feature querying. The query is written by hand	Not discussed
(Haider, 2012)	Ontology is created manually based on product reviews for certain mobile models	Opinion polarization score is calculated manually, from a set of reviews.	Not discussed	Not discussed
(Freitas and Vieira, 2013)	Existing movie ontology and hotel ontology is utilized	Associated polarity is calculated in real-time using OpLexicon	Not discussed	Not discussed
(Alkadri and ElKorany, 2016)	Existing hotel ontology HONTOLOGY is transalated into Arabic and further extended	Associated polarity is assigned by expert annotators by hand	Not discussed	Not discussed
(Sam and Chatwin, 2015)	Two ontologies are created manualy. Emotion Ontology is used for listing emotion words and Elecronic Product Ontology is designed for electronic product reviews	Using Ngram and Emotion ontology, sentiment analysis is performed	Not discussed	Not discussed
This Work	Ontology is created manually based on Face book users' comments/posts	Associated polarity is being calculated using SentiStrength tool and stored in Ontology	SPARQL for Protégé is used for feature querying. The query is automatically generated behind the formal user interface with the use of a mapping database	Recommendations are made using the previously calculated sentiment score

Facebook where the required object is found.

4. Before polarity calculation, text is processed to remove special characters(hashtag,@) and web URLs from the text. User comments are stored in a separate file for each school branch to apply sentiment analysis.
5. SentiStrength API is used to find out the polarity score of each user comment on an eleven point scale ranging from -5 to 5 and update the polarity table as shown in Table 2.
6. Using a polarity table calculate an overall polarity value for the object. The calculated sentiment score is saved back into the ontology on object level and class level for future referencing.

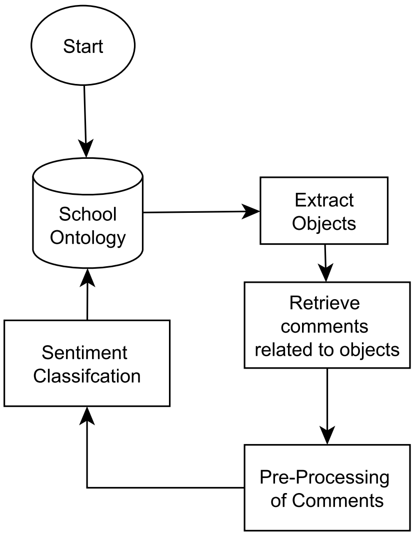

Figure 1: Proposed Approach.

For generating a recommendation request, a formal application user interface is designed that queries the ontology with the help of a mapping database. We use SPARQL query for the retrieval of desired classes and objects that fulfills user request from an ontology. Each selected instance object is compared with others, taking into account class level and object level sentiment scores. As a result, a valid recommendation is generated.

4 METHODOLOGY

4.1 Ontology Engineering

The school ontology is created manually using Protégé. School is the base entity/class for all the schools. All the schools are subclasses of the school entity/class. Whereas respective branches of each school are the real instances of any subclass of school. This has been done considering each branch of one school may have different Locations, GenderwiseOrientation like Boys Campus, Girls Campus etc. Similarly it is possible that one branch is primary and the other is secondary and like.

In Figure 2. an abstract school ontology is depicted containing fewer schools and their branches. These schools and their branches are the ones which are mostly discussed on the target forum and for which sentiments are mostly expressed. The names of schools are replaced with generic names for objective reasons. One may note that the data field SentiScoreMain relates to the sentiment score of a Class(School), however, SentiScore data field is for

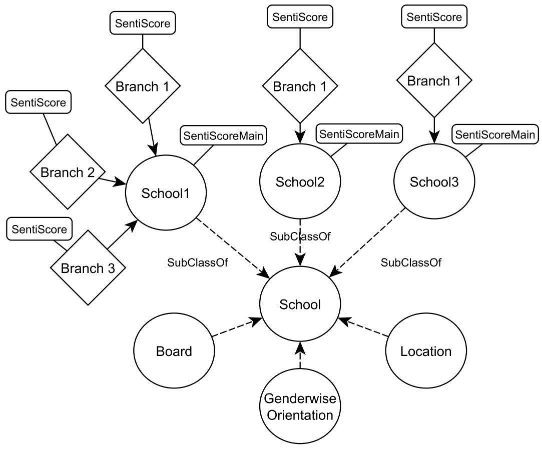

Figure 2: School Ontology.

storing the same for an object(School Branch).

4.2 Sentiment Analysis

JENA (a JAVA based API for ontology manipulation), is utilized to fetch all the relevant classes(school names) and their objects (School Branches) from the School ontology for which comments are to be retrieved from the Facebook. Facebook4J is utilized to fetch first *m* comments from Facebook. These comments are stored in a comments file. For each object o, randomly *n* subjective sentences are extracted from the comments list into a separate file manually, where object *o* is the target of opinion.

On completion of pre-processing phase, resulting comments are submitted to SentiStrength (another JAVA API for measuring sentiments on the scale of -5 to 5). Each sentiment related to an object(School Branch) is stored at object level and the class(School) is stored on class level. Algorithm 1 details the logic for generation of polarity table which will be utilized for Sentiment score calculation. Here, polarities corresponding to each object are being stored in a table row.

4.3 Polarity Calculation

Given the polarity row for each object, a sentiment score is the weighted average of all the polarity counts stored against each polarity value for a given object. Equation(1) will be used to calculate the sentiment score $F(o)$ for an object o.

$$F(o) = \sum_{z=-5}^{5} (z * Polarity(o,z))/n \qquad (1)$$

Where n=Total no of opinionated comments for object o and Polarity(o,z)is the frequency of Z polarity

Algorithm 1: Polarity table generation.

```
Input: Ontology with Classes(E),
       Attributes(A) and Objects(O)
Parameters: Empty set of Facebook
            Comments (C)
Output: Sentiment polarity table P with each
        row corresponds to each object o ∈ O
        and having eleven columns
        corresponding to each polarity value
        z ∈ Z (-5 to 5)
foreach e ∈ E which has a direct object do
    if e has object then
        foreach o ∈ O which is instance of
        Class e do
            //Retrieve m comments from
            Facebook
            C ← RetrieveComments(m)
            //Copy first n comments to Array S
            having object o in it and contains
            some user's sentiment
            S ← CopyComments(n)
            Insert a row in P for object o and
            initialize cell values to 0
            foreach s ∈ S do
                r ← GetPolarity(s)
                P(o,z) = P(o,z) + 1 ; such that
                r = z
            end
        end
    end
end
return P;
```

value for object o in given set of comments where object o is the subject of opinion. It should be noted here that value of n remains the same for all the objects to fairly calculate the sentiment score.

$$G = \sum_{i=1}^{n} F(o_i)/n \tag{2}$$

Similarly, in a scenario where sentiments relating to objects(School Branches) are expressed only, Equation(2) can be used to find an average sentiment score G for any class (School) itself, where $o_1, o_2, \ldots, o_n$ are objects of that class.

5 EXPERIMENTAL RESULTS

We have considered four school branches for experimental purpose. For each school branch, we extracted randomly 250 opinionated comments from the Facebook group. A separate file was created for each school branch and each file contains comments separated by newline character. Each file is then processed through SentiStrength tool, hence, polarity table is generated. Each row of the table is associated with an individual School branch. The resulting Polarity table is shown in Table 2.

Table 2: Polarity table.

School (O)	Polarity											
	-5	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**Polarity Score**
School1- Branch 1	0	2	3	4	10	1	25	33	38	54	80	3.144
School1-Branch 2	1	2	5	7	5	0	29	40	44	51	66	2.912
School2 -Branch 1	8	4	34	48	52	3	36	27	20	13	5	-0.316
School3-Branch 1	0	1	4	8	6	1	40	33	41	66	50	2.82

If a user is interested in all school on a specific location A, then the comparison will be made among all the school branches which are located at A and the one with the highest sentiment score will be recommended.

Table 3: School-level Polarity Table.

School (O)	Calculation	Polarity
School 1	(Sentiment Score (Branch 1)+ Sentiment Score (Branch 2))/2	(3.144+2.912)/2 = 3.028
School 2	Sentiment Score (Branch 1)	-0.316
School 3	Sentiment Score (Branch 1)	2.82

Similarly, if a user is asking that which school is better than the other, then the overall sentiment score of the school will be taken into account. Table 3 shows the calculation for different schools.

6 CONCLUSION AND FUTURE WORK

These days recommender companies are moving towards the semantic-based approaches for developing efficient recommendation systems (Bernstein et al., 2016). This motivated us to utilize an ontology for knowledge domain modeling and to perform sentimental analysis on user-generated reviews to develop a recommendation system. In the proposed system, calculated sentiment scores are stored in the ontology. This characteristic distinguishes it from other systems where scores are not saved anywhere. The scope of our future work involves considering Facebook likes into the calculation of sentiment score. Additionally, we look forward to implementing feature-level sentiment analysis between two schools and automate the task of subjectivity classification.

REFERENCES

Alkadri, A. M. and ElKorany, A. M. (2016). Semantic feature based arabic opinion mining using ontology. *International Journal of Advanced Computer Science and Applications*, 7(5):577–583.

Bernstein, A., Hendler, J., and Noy, N. (2016). A new look at the semantic web. *Commun. ACM*, 59(9):35–37.

Binali, H., Potdar, V., and Wu, C. (2009). A state of the art opinion mining and its application domains. In *Industrial Technology, 2009. ICIT 2009. IEEE International Conference on*, pages 1–6. IEEE.

Freitas, L. A. and Vieira, R. (2013). Ontology based feature level opinion mining for portuguese reviews. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 367–370. ACM.

Haider, S. Z. (2012). An ontology based sentiment analysis: A case study.

Kontopoulos, E., Berberidis, C., Dergiades, T., and Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert systems with applications*, 40(10):4065–4074.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Noy, N. F., McGuinness, D. L., et al. (2001). Ontology development 101: A guide to creating your first ontology.

Sam, K. M. and Chatwin, C. (2015). Ontology-based sentiment analysis model of customer reviews for electronic products. In *Encyclopedia of Information Science and Technology, Third Edition*, pages 892–904. IGI Global.

Thakor, P. and Sasi, S. (2015). Ontology-based sentiment analysis process for social media content. *Procedia Computer Science*, 53:199–207.

Yaakub, M. R., Li, Y., Algarni, A., and Peng, B. (2012). Integration of opinion into customer analysis model. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03*, pages 164–168. IEEE Computer Society.