International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-3 Issue-5 September-2012

Swing Detection Using Onto Tree

Rakesh Pandit[1], Parag Sohoni[2], Sachin Patel[3]
Patel College of Science and Technology, Indore, India[1,2,3]
Asst. Prof-IT[1], Mtech-IT 4th Sem[2], HOD-Asst.Prof-IT[3]

Abstract

In this paper we present an hybrid approach for Ontology designing by combining Data structure storage techniques and Ontology characteristics to improve understanding about a passages/statement or opinion. By applying classification task we focus on forming a right positive and left negative sub tree and to calculate its positive or negative swing. It serves many Purposes such as, e.g., opinion mining, Inclination and proper designing. Most Existing techniques utilize a list of opinion (also called opinion Lexicon). Opinion words are of both Characteristics (positive or negative) .By using binary search technique we have developed Unique onto trees that utilize tree structure to improve feature ranking. In the end we have formed a swing table that projects the inclination of any statement/opinion or passage.

Keywords

Ontology, Onto Tree, Information Science, Data Structure, Opinion mining, Feature Ranking,BSOT, RPOT, LPOT, Swing Table

1. Introduction

In recent years ontology has emerged as an effective tool in the field of information science and it also affectively resolve complex problems like storage and sorting Of it. It has also been well discussed that we can form ontology as per ones; Understanding. Increased understanding about a given opinion or written statement will surely be helpful in so many scenarios from market research to product launching, feature ranking and most importantly it will allow all to prepare and will help us to reduce cost with a serious margin. Common word study out of any statement will surely help us to save time of all from a manufacturer to customer / buyer to purchaser/artist to admirer and writer to reader.

Any popular movie/book or any other product can get hundreds of reviews which will be the combination of all types from good to bad or average as well. Some opinions are also in the form of paragraph and which makes it hard for a potential reader to read them to make an informed decision on whether to follow that certain reviews or not . If read views are only full of admiration and in reality product is not that much good enough then any reader will surely loose his/her money and time, to go through long ,complex and misguided review. Ontology has been studied by many researchers in recent years. Although some methodologies for building ontologies have been proposed to improve the ontology development process [10, 11, 12, 13, and 14] building ontology manually is a time-consuming and laborious activity that requires the work of highly trained ontology engineers. Moreover for designing ontology there is no pre-defined thumb rule. Like any other paper or approach, it also has some bottlenecks, the problem with this approach is that it has limited range of pre declared words; either we can say there is no specific way to know the orientation of any word. In our paper we have designed an ontology creation by passing and checking any statement/comment or paragraph ,our design of BSOT will read the complete sentence in a shot and will produce and categorize all possible adjectives/sentiments/emotions and special purpose words and will also produce the +ve and –ve shades in its respective left and right.

First of all By checking and finding its +ve and –ve shades we can specially find swing of any statement/review /comments or any long sized paragraph .instead of observing the current sentence alone it can read and find out the swing inclination a complete paragraph or a set of paragraphs .

Secondly it is based on our Results, we are quite confident to say that when there are multiple conflicting opinion words in a sentence, existing methods are unable to deal with them well. The results show that the new method outperforms the existing methods significantly.

We have organized our discussion as follows. In section 2 we discussed previously work done in the field of swing detection. Under which we also discussed our method and the evaluation too. We also discussed about Sentiment analysis and web2.0.In

section 3 we described the problem definition. in section we described about our proposed technique, which comprises of the concept under taken, the types of the onto tree, our proposed algorithm and the pattern rules to create onto tree. In section 6 we conclude with a summary of what we have achieved and in section 7 future works is given.

2. Related Work

This section provides an overview of the relevant related work done in the area of the ontology. Ontology has been studied by many researchers in recent years. Although some methodologies for building ontologies have been proposed to improve the ontology development process [10, 11, 12, and 13], building ontologies manually is a time-consuming and laborious activity that requires the work of highly trained ontology engineers. Moreover, an ontology that is built manually tends to be biased toward its developer's view. Typically, ontologies can be generated from various data types such as textual data, dictionaries, knowledge bases, semi structured schemata, and relational schemata. Most works related to automatic ontology construction have been directed toward extracting ontology from texts [14]. A typical approach in ontology extraction from text first involves term extraction from a domain-specific corpus at hand.[1] proposed an approach to refine the merged ontology, which tailors the merged ontology Through removing the instantiation redundancy and subsumption redundancy and checking the consistency of the consequential ontology. In [2] a framework for designing information systems on top of OWL ontologies has been proposed. In [3] a machine-made system of ontology construction was brought up. In [5] construct a two-tier ontology. In [6] a new method for learning ontologies has been developed and this method applied LDA model to extract topics as the concept of ontologies from given document corpus. Paper [7] presented a word-based method for extracting sentiment from texts. Paper [8] proposed an effective method for identifying semantic orientations of opinions expressed. In[9] a method for determining a sentiment associated with an entity has been proposed. Our approach of detecting swing by using an noto tree mainly focusses on finding the inclination and arranging a given ontology by forming an onto tree.Thge reason behind using tree is not only because it properly represents any informationb but also it allows to traverse and search both in upward/downward and left /right sides as well. Swing table used in our method helps to project the complex detail in a simple and systematic manner .

2.1. Other methods

A basic task in swing analysis is classifying the inclination of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry," "sad," and "happy." Early work in that area includes Turney and Pang who applied different methods for detecting the polarity of any statement/opinion and movie reviews even a lengthy one respectively. This work is at the document level. It is also possible to scale all this swing at a multilevel scale as well, which was attempted by Pang and Snyder (among others) expanded the basic task of classifying a movie review as either positive or negative to predicting star ratings on either a 3 or a 4 star scale, while Snyder performed an in-depth analysis of restaurant reviews, predicting ratings for various aspects of the given restaurant, such as the food and atmosphere (on a five-star scale).

Scaling system is also an approach for analyzing swing, the use of a scaling system whereby words commonly associated with having a negative, neutral or positive swing with them are given an associated number on a -5 to +5 scale (most negative up to most positive) and when a piece of unstructured text is analyzed using natural language processing, the subsequent concepts are analyzed for an understanding of these words and how they relate to the concept. Each concept is then given a score based on the way sentiment words relate to the concept, and their associated score. This allows movement to a more sophisticated understanding of sentiment based on an 11 point scale. Alternatively, texts can be ranked with positive and negative keyword ranking score if the goal is to determine the sentiment in a text rather than the overall polarity and strength of the text. Another research direction is subjectivity/objectivity identification. This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective. This problem can sometimes be more difficult than polarity classification the subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (e.g., a news article quoting people's opinions). Moreover, as mentioned by Su, results are largely dependent on the definition of subjectivity

used when annotating texts. However, Pang showed that removing objective sentences from a document before classifying its polarity helped improve performance.

The more fine-grained analysis model is called the feature/aspect-based swing analysis. It refers to determining the opinions or swings expressed on different features or aspects of entities, e.g., of a cell phone, a digital camera, or a bank. A feature or aspect is an attribute or component of an entity, e.g., the screen of a cell phone, or the picture quality of a camera. This problem involves several sub-problems, e.g., identifying relevant entities, extracting their features/aspects, and determining whether an opinion expressed on each feature/aspect is positive, negative or neutral.

Computers can perform automated swing analysis of digital texts, using elements from machine learning such as latent semantic analysis, support vector machine, "bag of words" and Semantic Orientation Point wise Mutual Information .More sophisticated methods try to detect the holder of a swing (i.e. the person who maintains that affective state) and the target (i.e. the entity about which the affect is felt). To mine the opinion in context and get the feature which has been opinionated, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text.

Open source software tools deploy machine learning, statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, including web pages, online news, internet discussion groups, online reviews, web blogs, and social media.

2.2. Evaluation

The accuracy of a swing analysis system is, in principle, how well it agrees with human judgments. This is usually measured by precision and recall. However, human raters typically agree about 70% of the time. Thus, a 70% accurate program is doing as well as humans, even though such accuracy may not sound impressive. If a program were "right" 100% of the time, humans would still disagree with it about 30% of the time, since they disagree that much about any answer. More sophisticated measures can be applied, but evaluation of swing analysis systems remains a complex matter. For swing analysis tasks returning a scale rather than a binary judgment, correlation is a better measure than precision because it takes into account how close the predicted value is to the target value. Swing analysis was used to test the relationship between Internet financial message boards and the behavior of the stock market to find a strong correlation between posts and volume of stock.

2.3. Sentiment analysis and Web 2.0

The rise of social media such as blogs and social networks has fueled interest in sentiment analysis. With the proliferation of reviews, ratings, recommendations and other forms of online expression, online opinion has turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities and manage their reputations. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and auctioning it appropriately, many are now looking to the field of sentiment analysis. If web 2.0 was all about democratizing publishing, then the next stage of the web may well be based on democratizing data mining of all the content that is getting published.

One step towards this aim is accomplished in research. Several research teams in universities around the world currently focus on understanding the dynamics of sentiment in e-communities through sentiment analysis. The Cyber Emotions project, for instance, recently identified the role of negative emotions in driving social networks discussions. Sentiment analysis could therefore help understand why certain e-communities die or fade away (e.g., MySpace) while others seem to grow without limits (e.g., Face book).

The problem is that most sentiment analysis algorithms use simple terms to express sentiment about a product or service. However, cultural factors, linguistic nuances and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment. The fact that humans often disagree on the sentiment of text illustrates how big a task it is for computers to get this right. The shorter the string of text, the harder it becomes.

3. Problem Definition

Each feature in any ontology has its specific importance for information representation .The correlated study of such features can project their differences and importance. Instances are usually objects under concepts, which may partially or completely reflects any ontology for its better

understanding we should try to find there inclination on both positive and negative scale. Therefore, instances can express concepts more clearly. Ontology is convenient for information sharing, establishing the network information under a well formatted structure which allows information sharing and knowledge reuse, and describing online documents through ontology description languages. Traditionally, ontology construction usually uses the following methods for ontology construction: relational analysis, clustering, and formal concept analysis. The relational analysis (FCA) discovers and clusters the relations of keywords, such as synonyms, roots, hypernyms, and hyponyms. It then constructs the ontology by manmade or other methods. For example, in the framework of WebOntEx (Web Ontology Extraction) [15] the keywords in the web pages of particular domains are clustered by relations, such as synonym, in the WorldNet to obtain concepts. The ontology is constructed by means of the positions relevant to the HTML tree formed by the connection of the web pages in the tag information and particular domains. J. U. Kietz used the relations between hypernyms and hyponyms in the GermaNet to acquire concepts. Clustering usually groups keywords of documents into clusters and constructs ontology by selecting representative concepts from each cluster. For example, TextOntoEx used the semantic roles to cluster keywords. Formal concept analysis uses the binary relation matrix between documents and vocabulary to generate the supremum concept set. The inter-conceptual hierarchical relation and a complete partial sort is formed by the sets of all the concepts. After constructing the inter-concept level relation by means of FCA construction, the concept figure of ontology is thus constructed.

4. Proposed Method

The explanation for using a tree structure to represent an ontology can be very well understand by observing the fact that, first of all there is no such specific thumb rule defined for representing any ontology ***"An ontology can be represent as per one's understanding and*** *requirement ".*

4.1. Concept Undertaken

The logic which we have adopted for designing our ontology follows a very famous storage scheme under data structure called trees.
Example -

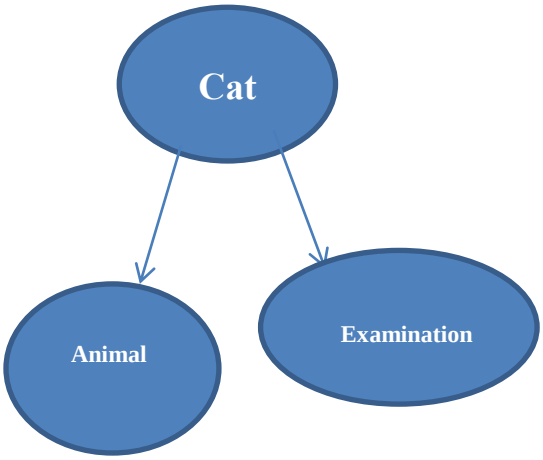

Figure 1: Ontology 1

The tree created above represents that an ontology can also be formed or designed in such a way, in the e.g. Depicted above shows that "Cat "A word, Animal or an Examination type can be arranged in such a manner" ,Thus by using the same further we have formed different type of onto trees to represent all possible sentiments and which finally leads to find its swing or inclination.

4.2. Type of Onto Tree

NOT –NULL –ONTO- TREE – An onto tree with no child at all.
RPOT- This ONTOTREE will contain all positive sentiment at its right side.
LPOT- This onto tree will contain all negative emotions at its left side.
BSOT- It will follow a proper Binary search tree approach

4.3. Algorithm

Proposed Algorithm

1. **Algorithm Beginning**
2. **Step 1**- Create a user defined data type (Struct), having attributes
3. **Struct node --> attributes all**
4. **Struct node *left;** /* Self-referential pointer variable
5. **Char array [ ];** /* used to store data
6. **Struct node *right;** /* Self-referential pointer variable

Step 2- We have used 4 different Function Prototype to solve different purpose.

6. Insert ( ) /* It is used to enter the string from main body

7. Check ( ) /* It matches the words from given string entered
8. Display( ) /* It is used to display tree formed
9. Main( ) /* It comes under main body of the program and performs all Logics
Step 3-Insert string and search it in the Insert function
10. Now it creates Root node in them Apply check function for matching the (positive and negative) sentiments in the array.
11. It generates / Creates left and right Nodes Referential array ends
/* Functionality of Check function

12. Input [i] = array of input string
13. Post [ ] = array of the +ve sentiments
14. Negt [ ] = array of the –ve sentiments
15. Match 1st word of Input [i] with both array

16. If found
If Input [i]<- Post [ ]

17. Then generate Left Node

18. Left <- Input [i]

19. End if

20. Input [i] <- Negt [ ]

21. Then generate right node
22. Right <- Input [i]

23. Else
24. i = i +1 ;
25. **Step 4 –** Repeat Step 1
26. **Step 5-** Referential array ends
27. **Step 6 –** Display -> Display the sentiments
28. **Step 7 –** Print the tree in the input f

4.4. Pattern Rules to Create ONTO TREE

This is the pattern how we have categories the swing of statement-

Rule 1- If (Positive value)
Create Left Node
Rule 2- If (Negative value)
Create Right Node
In some exceptional cases like "Work" is a neutral verb.
The system detects the patterns and
Then applies the rules written above.

Based on this rules we are creating an onto tree for each given statement.
Constructed ONTO TREE representing Swing is as Follows-

String 1- Ram is a good and nice boy.
For such a string our method suggests that
It only contains positive words and as per our algorithm only left root will be formed

Onto Tree on the basis of Algorithm is as follows:

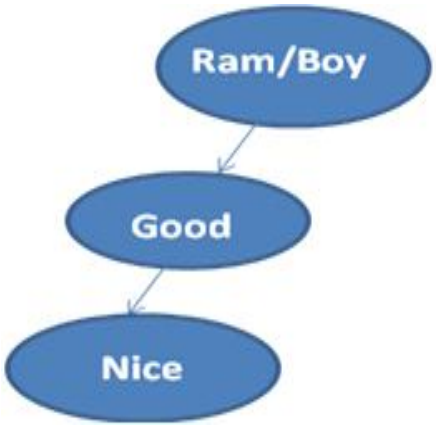

Figure 2: Ontology 2

String 2. John got failed in examination, that's why he is sad.

For such a string our method suggests that It only contains Negative word and only right roots will be formed.

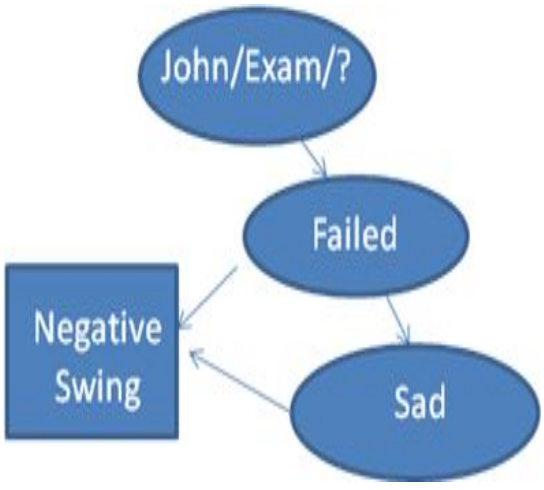

Figure 3: Ontology 3

String3- I have seen a movie yesterday in which music was outstanding but acting was bad

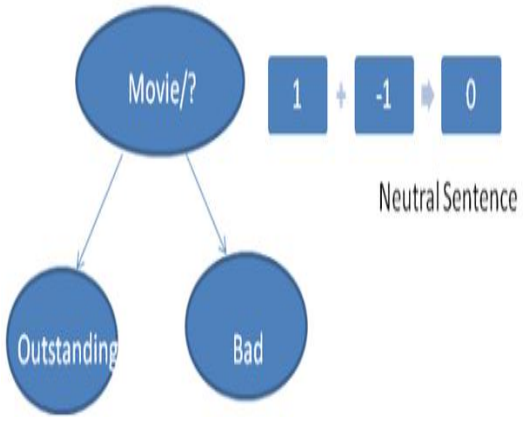

Figure 4: Ontology 4

Table 1:Swing table to detect Inclination

String No	Root Node values	Left Sub Tree values	Right Sub Tree Values	String Inclination
S1	Ram/Boy Or Empty Or Complete String	Good ,Nice	Bad	Positive
S3	John/Examination Or Empty Or Complete String	-	Failed ,Sad	Negative
S4	Movie Or Empty Or Complete String	Outstanding	Bad	Neutral/Contradictory

5. Conclusion

This paper proposed an effective method for identifying swing and to make an onto tree from a given statement or a string. It resolves an important purpose of finding a swing inclination of any string and also helps to arrange a high amount of data records by forming Onto-tree .Previous research and paper work have used a long and time consuming approaches and more importantly they were unable to produce any well-formed and easy to understand structure. Our method uses the concept of data structure as well which makes it more complete. Tree and Table based results shows that the proposed method performs better.

There are adjectives that have no orientations. There are also many words whose positive or negative swing depend on contexts in which they appear. We will work to deal with this problem.

References

[1] Jingyi Zhang, Yanhui Lv.” An Approach of Refining the Merged Ontology”. 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012).

[2] Petr K˘remen and Zden˘ek Kouba.” Ontology-Driven Information System Design”. IEEE Transactions on Systems, MAN, and Cybernetics—PART C: Applications and Reviews, VOL. 42, NO. 3, MAY 2012.

[3] Yao-Tang Yu, Chien.”Chang Hsu.” A Structured Ontology Construction By Using Data Clustering And Pattern Tree Mining”. Proceedings of the 2011 International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011.

[4] Choi, Key-Sun. "IT ontology and semantic technology." In Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference on, pp. 14-15. IEEE, 2007.

[5] Mei-ying Jia, Bing-ru Yang-“Research on Domain Ontology Construction in Military Intelligence”. 2009 Third International Symposium on Intelligent Information Technology Application.

[6] Zhijie Lin, Rui Lu, Yun Xiong, Yangyong Zhu.” Learning ontology automatically using topic model”. 2012 International Conference on Biomedical Engineering and Biotechnology.

[7] Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. "Lexicon-based methods for sentiment analysis." Computational linguistics 37, no. 2 (2011): 267-307.

[8] Xiaowen Ding, Bing Liu, Philip S. Yu.” A Holistic Lexicon-Based Approach to Opinion Mining”. WSDM’08, February 11-12, 2008, Palo Alto, California, USA.

[9] Namrata Godbole, Steven Skiena, Manjunath Srinivasaiah.”Large Scale Sentiment Analysis”.US 2008/0270116 A1.Oct.30, 2008.

[10] O.Corcho, M. Fernandez, A. Gomez-Perez, Methodologies, tools and languages for building ontologies. Where is the meeting point? Data & Knowledge Engineering 46 (2003) 41–64.

[11] A. De Nicola, M. Missikoff, R. Navigli, A software engineering approach to ontology building, Information Systems 34 (2) (2009) 258–275.

[12] M. Fernandez, A. Gomez-Perez, N. Juristo, and Methontology: from ontological art to ontological engineering, Workshop on Ontological

Engineering. Spring Symposium Series (AAAI '97) Stanford, USA, 1997.
[13] Y. Sure, S. Staab, R. Studer, On-To-Knowledge methodology (OTKM), in: Steffen Staab, Rudi Studer (Eds.), The Handbook on Ontologies, 2004.
[14] C. Tao, D. Embley, Automatic hidden-web table interpretation, conceptualization, and semantic annotation, Data & Knowledge Engineering 68 (2009) 683–703.
[15] K. W. Tan, H. Han, and R. Elmasri, "Web data cleansing and preparation for ontology extraction using WordNet," Proceeding of the First international Conference on Web information systems engineering, vol. 2, pp. 11-18, 2000.