A NOVEL METHOD OF OPINION EXTRACTION FOR PRODUCT OPINIONS

Ramandeep Sandhu (Sr. Lec.)
Department of Computer Science & Engineering, CGC Technical Campus Jhanjeri (Chandigarh), Punjab, India
raman8deep@gmail.com

Kiran Jyoti (AP)
Department of Computer Science & Engineering, Guru Nanak Dev Engineering College (Ludhiana) Punjab, India
kiranjyotibains9@gmail.com

ABSTRACT

An opinion is attached with emotions and in more wits its interpretation of facts. “Opinions” are key influencers of someone’s behavior. For a product, it’s a thinking that leads to a decision, where a product stands. Through opinion mining, we can know about contents of whole product opinions. Opinion mining (sentiment analysis) is a challenge where opinions mean less efforts and best results to correct information about a product. Opinions are not a dissertation which can be left unread but it’s a track to reach in mind of people to get views about a product. This paper presents an algorithm to extract opinions from a micro blogging site on which an organization trusts and consider its reviews. By this technique, database of online opinions can be created by far.

KEYWORDS

Opinion mining, product opinions, opinion extraction

1. INTRODUCTION

Opinion is one of the important types of information on the web which expresses attitude as user words with respect to some topic. An individual needs opinions posted by others when: want to purchase a new product (mobile) and has a big question in mind that How to Select it? Businesses, organizations, individuals all need an effective opinion mining system. User desires for purchasing a new product surrounds a bag of questionnaires as shown in Figure 1.

Web is a great place for people to express their opinions [1]. People feel free to say about a product which they are using or used and want to say about it to others. This topic is basically concerned with our daily life activities. What we want, what we purchase, what we think before purchasing a new product [2], how we evaluate product features etc. Public opinions are every bit important for suppliers as well as consumers; it gives people the chance to correlate their ideas together. Opinions are always analyzed to eliminate noisy and polluted type of opinions.
Figure 1: User’s need of opinion mining system

Through the Opinion Mining [10], we can know about contents of whole product reviews and also about reputation [4, 5] of your own or competitor’s product. Opinions on the web can be accessed with fast access of net. It’s possible only with World Wide Web. Opinions should not be raw data. It should be topic-focused means what works for one subject should not work for another subject. Example shows, if topic is “mobile” then what a person can express for mobile should not express for “politics” topic.

As opinions not resides in air. These are linked to a domain probably. Domain is a boundary which disconnects one kind of opinions from another. Opinion can be a single sentence, a document or many sentences or many documents. It can now be found almost everywhere blogs, social networking sites.

**An opinion is a quintuple:**

![Figure 2: Opinion a quintuple](image)

An entity is an object about which opinions are posted. The opinion orientation/polarity [6] can be categorized as:

Positive: Enhances the reputation of a product.
Negative: Collect deficits of a product.

Opinion holder is the governor of an opinion and normally refers to a person, a state or an organization as well as the corresponding pronouns. Entity is also called object. Aspect is also called feature, attribute, facet, etc. Opinion holder is also called opinion source. “Opinions” are key influencers of our behaviors. Our beliefs and perceptions of reality are conditioned on how others see the world. Whenever we need to make a decision, we often seek out the opinions of others. In the past, Individuals seek reviews from friends and family. Organizations use surveys, focus groups, opinion polls, consultants. Word-of-mouth on the web is personal experiences and opinions about anything in reviews, forums, blogs, Twitter [8], micro-blogs, etc.

Opinions are words-of-mouth and it can’t be formality to say something about an entity, an ordinary person gives his views after thinking and evaluating a product [2]. Opinions [6, 7] are always valuable, but how valuable it remains after extraction is main venture of opinion extractor. As people spend their time on web in posting opinions because their thinking impels them to help others in becoming more familiar with the new products launched. It’s also a case that sometimes, a native person is unable to grasp from help of a product, which he want to purchase but feel free to take decision after reading posted opinions by others. Opinions are not helping people in one means many open conducts its providing to people which are helping them to take self-decisions in life and to encourage with new products launched. As web is the platform where person can post and review opinions. It acts as an interface passing information in both directions.

Opinion mining an emerging area of research is a subtopic of information retrieval (IR) with considerable research done. It’s concerned with applying computational methods to the treatment of subjectivity in text, with a number of applications in fields such as recommendation systems, contextual advertising and business intelligence. Several methods exist to determine an author’s view on a topic from natural language textual information. These generally employ some form of machine learning approach, and have varying degrees of effectiveness.

Opinion mining is used for tracking the frame of mind of the public about a particular product. It identifies the people arguments/ judgments about a product in a text.

2. RELATED WORK

Researchers have performed many kinds of studies on opinion extraction and summarization on web [3]. This section conceptualizes the previous published work. The focal point is what to do with posted opinions and how to gather them, also web is considered as a source of opinion mining.

Ramandeep Sandhu and Rahul Mehta (2011) in paper entitled “Applying Opinion Mining to Organize Web Opinions” proposes an effective system to organize web opinions and designed a system to organize web opinions at the time when user is posting, before actually being extracted by expertise. The new system (Opinion Organization System) provided four stages. In first stage, it provided a list of several product categories and user selected at least one. In second stage, a list of selected product relevant features is displayed to the user. In third stage, user firstly selected features for which wants to express opinions, then used polarity based P set and N set containing adjective words list and in fourth stage, used thumb selection table to add opinions.

Aymen Elkhlifi et. al. (2010) in paper entitled “Automatic Extraction and Classification Approach of Opinions in Texts” implements a system named sec-op (a system of extraction and classification of opinions) in java language based on semantics. The system works in four
modules. It uses SEG-SEN segment text for POS tags, uses the API of tree tagger, wordnet for synonyms and HSO & LIN to complete term similarities. To relate lexical semantics and subjectivity, SemEval2007 corpus is used. The result indicates superiority of participated system to SemEval2007.

Alexander Pak and Patrick Paroubek (2011) in paper entitled “Twitter as a Corpus for Sentiment Analysis and Opinion Mining” describes Twitter as a Corpus for opinion mining. They used twitter, the most popular micro blogging platform for opinion mining. Also automatic corpus was collected and then linguistic analysis of the collected corpus war performed. Opinion classifier is built to find positive, negative and neutral opinions/feelings for a document.

Sheng Feng el. al. (2010) in paper entitled “Recommended or Not Recommended? Review Classification through Opinion Extraction” outlines Review Classification through Opinion Extraction for online product reviews. The opinions are classified to judge whether recommending the product for which posted or not? The work is performed in two steps. First step involved mining the product features and relevant opinions. Second step involved calculating the opinion sentence orientation based on polarity. Also opinion strength is evaluated for each sentence.

SentiWordNet dictionary was used to generate opinion words. POS tagging was performed to mine product features. Then pair named “Identifying Feature-Opinion Pairs” from product features and opinion Words were generated. This pairing was used to determine polarity and opinion strength to generate message “recommended/not recommended” by aggregating feature-opinion pairs.

3. THE PROPOSED SYSTEM

Proposed Roadmap of Opinion:

An opinion can be described as seven tuples as shown in Figure 3.

![Figure 3: Aspects of an Opinion](image)

Opinion words can be categorized as two types:

Positive opinion words
Negative opinion words
Also Ranking of a product can be categorized as two types because people believe in ranking system of products.

Positive ranking
Negative ranking

**Positive opinion words:** The words which express positive significance and the sense in which these words are used basically enhance the value of the product in the market. As all in today’s world are after plus points of a product. Plus points are known as positive values. Positive ranking can be further segregated in two types: high positive ranking and low positive ranking.
Examples: good, ultimate, amazing, marvelous etc.

![High positive rank](image)
![Low positive rank](image)

**Figure 4: Star allotment for positive rank of a product**

**Negative opinion words:** The words which express negative significance and the sense in which these words are used basically disgracing the value of the product in the market. When a company launches a new product, it’s not always possible that the product is completely satisfactory. Until the product is tested by customer for whom company is manufacturing, defects can’t be attributed completely. So its best way to capture the negative opinions and for this purpose, negative opinion words help. Also negative impact can be further segregated in two types as: high negative impact and low negative impact.

Examples: bad, poor, disgraceful, awful etc.

Micro blogging sites are social networking sites that enable its users to send and read text-based posts. Various types of micro blogging sites in today era are twitter, facebook, twitter etc. In our research, twitter is selected to extract opinions known as tweets.

Twitter [8, 9] enables its users to send and read text-based posts of up to 140 characters, informally known as "tweets". Twitter is an information network and communication mechanism that produces more than 200 million tweets a day. The Twitter platform offers access to that corpus of data, via our APIs. Each API represents a facet of Twitter, and allows developers to build upon and extend their applications in new and creative ways. It's important to note that the Twitter APIs are constantly evolving, and developing on the Twitter Platform is not a one-off event.

A tweet is a post or status update on Twitter, a micro blogging service. Because Twitter only allows messages of 140 characters or less, "tweet" is as much a play on the size of the message as it is on the audible similarity to Twitter.

Step 2: Collect data in terms of opinion sentences expressed by individuals based on some timeline using API's.

With the Evri API, you can automatically, cost effectively and in a fully scalable manner Analyze text. Get recommendations, discover relationships, mine facts and get popular data.
ALGORITHM FOR OPINION EXTRACTION FROM WEB:

Step 2.1: SELECT the query to get opinions for a particular product.
   E.g:- QUERY = "Nokia" (Nokia is a product)

Step 2.2: SEND the request to get the opinions for selected query from a particular website (e.g.
   TWITTER).
   REQUEST = http://search.twitter.com/search.json?q=.urlencode(QUERY)

Step 2.3: CONNECT to DB.

Step 2.4: WHILE end reaches
   Step 2.4.1: READ items as id, created_at, from_user, from_user_id, text, source, iso_language_code and query from web database
   Step 2.4.2: REMOVE special characters with the help of preg_replace function
      from the TEXT field.

Step 2.5: GET the iso_language_code = "india" & "england" from the database
   5.1: READ items in step 4 and INSERT into the local database

Step 2.6: CLOSE database connection.

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>bigint(20)</td>
</tr>
<tr>
<td>created_at</td>
<td>datetime</td>
</tr>
<tr>
<td>from_user</td>
<td>varchar(100)</td>
</tr>
<tr>
<td>from_user_id</td>
<td>bigint(20)</td>
</tr>
<tr>
<td>text</td>
<td>text</td>
</tr>
<tr>
<td>iso_language_code</td>
<td>varchar(10)</td>
</tr>
<tr>
<td>query</td>
<td>text</td>
</tr>
</tbody>
</table>

Figure 5: Fields of Extracted Opinions

Explanation of database fields:

1. id: It represents unique table id for each new opinion posted by user.
2. created_at: It represents time at which a particular opinion is posted on web (e.g. tweet
   posted on twitter).
3. from_user: It represents user name who writes opinions.
4. from_user_id: It represents the primary id of the user who posts opinions from his/her
   account.
5. text: It represents the opinion text. Each new tweet entry is new opinion posted by
   different id person at same or different time.
6. source: It represents the source from where opinions are generated means which web
   place a person uses to add opinions.
7. iso_language_code: It represents the country code to describe about the tweet.
8. Query: It indicates the name of the selected product.
CONCLUSION

This paper provides a novel method for online opinion extraction from twitter. Micro blogging sites are preferred rather than asking others for making decision to purchase any product. Twitter is a site which people access more in reviewing and sharing opinions with others. Huge database of online product opinions is created with useful record of each opinion. It also comes with result that until opinion extraction is performed, how it’s possible to judge high or low rank given by user for a product.

REFERENCES


Er. Ramandeep Sandhu is pursuing M. Tech (Thesis left) in Computer Science Engineering from Guru Nanak Dev Engg College, Ludhiana with agg. 80%. Her research interest is Web Mining and Opinion Mining. She has total 13 publications including National and International, published 01 paper in International Journal IJCSEA, presented and published 03 International Papers, 09 National papers and 02 National Seminars on Opinion Mining in Punjab. She has 4.5 yrs teaching experience- 1 year as Head of CSE department. Currently she is doing job in CGC Technical Campus, Jhanjeri (Mohali), Punjab as Sr. Lecturer in CSE department.