A NEW METHOD TO FIND SCORE VALUE FOR ONLINE OPINIONS

Ramandeep Sandhu (Sr. Lec.)

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

Kiran Jyoti (A.P.)

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

ABSTRACT

In today's world, we need information, not data to take a decision about a product. Opinions are data and become information after filtering. Also organizations products superiority is dependent on customer feedback of their products. Web is a platform providing facilities to post opinions for any product. A customer gives feedback when he uses a product and feedback is what customer feels after using it called sentiments/opinions. Opinions are subjective sentences and not objective as facts. People uses natural language to express sentiments and way of expressing each customer is different. Now it's inevitable in current and future era to do sentiment analysis called opinion mining. So that, what customer is saying about a product whether enhancing reputation or not, will be posted on web. Opinion Mining and scoring of opinions in fast competitive world is required as opinions posted for a trendy product can be in hundreds. If these opinions not analyzed properly and have been thrown as is it is on web then it's not possible for anyone to read all opinion sentences. So, a system is needed which automatically mines online opinions from web and score them (score value) properly correspond to positive and negative opinion words.

KEYWORDS

Opinion mining, product opinions, sentiment analysis, opinion analysis, opinion scoring.

1. INTRODUCTION

An opinion is attached with emotions and in more wits its interpretation of facts. "Opinions" are key influencers of someone's behavior means how a person evaluates something according to his/her thinking is an opinion. For a product, it's a thinking that leads to a decision, where a product stands. Through opinion mining [2], we can know about contents of whole product opinions [1]. Opinion can be unsubstantiated information as a subjective belief than fact-based belief and we need a system to mine opinions expressed in user generated content. In the world, one important kind of information is textual information which can be classified as either facts (always true) or opinions (can be true or false). Difference comes as facts are objective[Wikipedia] statements about anything and opinions are subjective[Wikipedia] statements which may be supported by an argument and reflects person's perspective, understanding, particular feelings, beliefs, and desires . Example shows difference in both as: Someone said that "Nokia is best mobile in the world" is an opinion as it can be proven true or can't be, but "I have a Nokia mobile" is a fact.
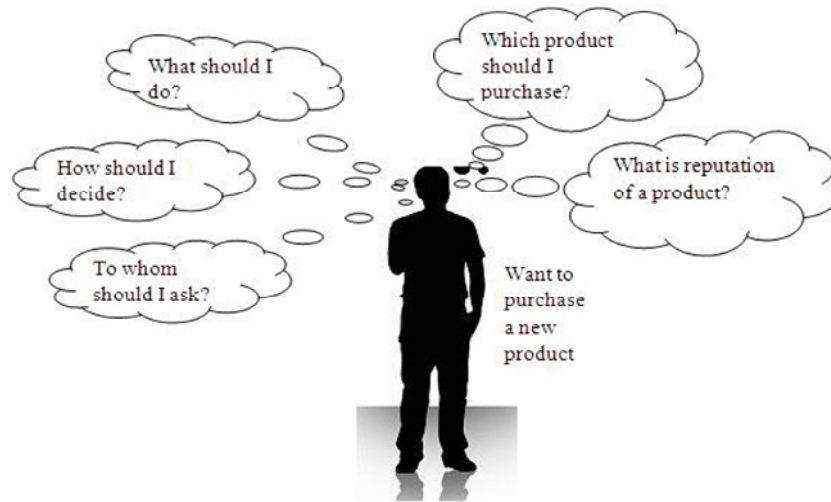
Figure 1: Importance of Product Opinions for a customer

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. Opinions are always valuable and important as shown in figure 1, 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 and it needs a system to manage opinions [4] at each glance.

2. RELATED WORK

Researchers have performed many kinds of studies on opinion extraction, decomposition 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.

Ramandeep Sandhu et. al. (2012) in paper entitled "Opinion Mining using Unequal Grouping for Product/Services Opinions" defined an opinion mining system helpful for both a customer and an organization. They applied this system for extracted dataset from web using Evri API

and after storing it in local database. They used an idea of Psychometric Scale known as Likert scale for rating different opinion sentences. Also a feature set was provided to do factorization of all feature included in opinionated sentences. During polarity(positive and negative) based [5] unequal classification of opinion sentences, labeled sheet helps which acts like final sheet of all mined 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.

3. THE PROPOSED SYSTEM

Proposed Roadmap of Opinion:

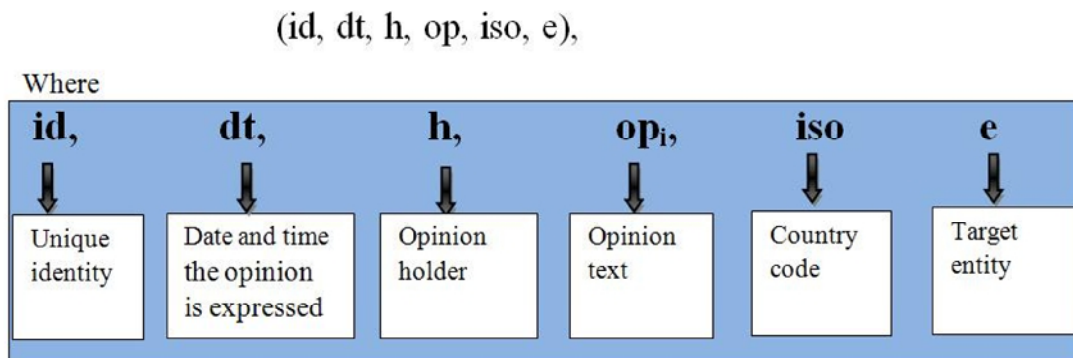An opinion can be described as six tuples as shown in Figure 2.

Figure 2: An Opinion and its aspects

Above identified six tuples is every bit important for each opinion posted as user who is posting opinions is not a single person and nor every time, same sentence has been posted. Id is unique id number for each new user, dt is date and time when a particular opinion was posted, h is holder means the person name who ha posted opinion, op is opinion sentence, iso is country code from where opinion was posted and e is entity means for which opinion has been posted
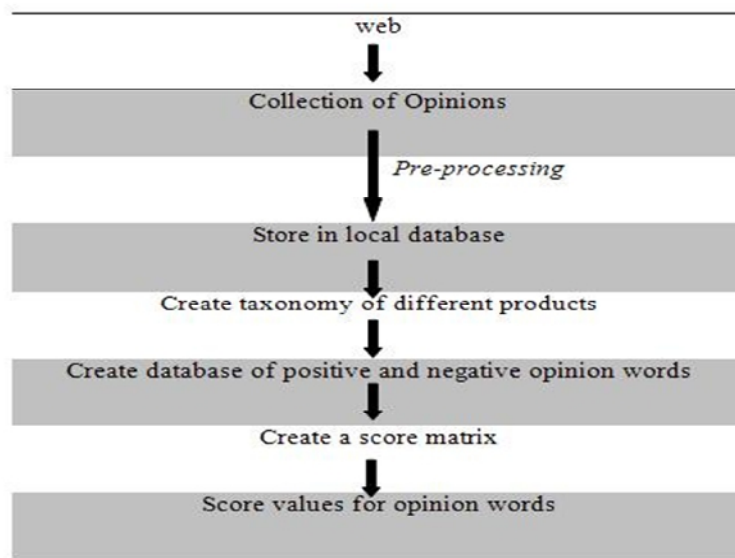
Methodology: Score Matrix for Product Opinions

Figure 3: Methodology for creating score matrix for a product

Explanation of different Phases:

Web: Web is social media + beyond

Twitter: A micro blogging site is an online social networking that 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. . In my research, twitter is selected to extract opinions known as tweets.

Word-of-mouth on the Web means personal experiences and opinions about anything in reviews, forums, blogs, Twitter, micro-blogs, etc. Comments about articles, issues, topics, review, etc.

- Postings at social networking sites, e.g., face book.
- Global scale: No longer – one's circle of friends
- Organization internal data
- Customer feedback from emails, call centers, etc.
- News and reports
- Opinions in news articles and commentaries

Identify micro blogging sites from where opinions can be extracted and which allow data access for mining web opinions using applications programming interface (APIs) as shown in figure 4.
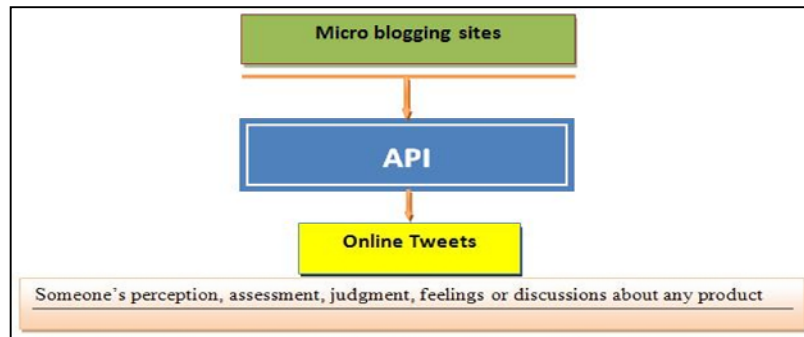
Figure 4: Online tweets extraction using API

Phase 1: Collection of opinions
Collect data in terms of online tweet sentences expressed by individuals based on some timeline using API's as shown in figure 4.
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 popularity data.
Phase 1.1: Perform de-noising (pre-processing)
Now a day's, web contains real world databases which are highly subject to noisy, missing and inconsistent data due to their usually enormous size data and their origin from multiple, heterogeneous sources.

Data pre-processing in data mining process is very important step. It is a component of web mining. Online opinions extraction [6] from web is named as data gathering phase in data mining. Online gathered data is raw data and never perfect as it always includes unwanted queries, not-domain oriented data etc.

In computer language, such kind of data is either redundant or irrelevant includes noisy and polluted data. So, knowledge discovery becomes very difficult. Data pre-processing includes cleaning, normalization, transformation and selection of data.

In this study, de-noising process is applied on extracted data deletes unwanted opinions e.g.
- Opinions which are not product relevant like if query selected is Nokia the all opinion sentences should be Nokia relevant and same for Samsung and Motorola products.
- Deleted blank opinion entries as empty record is polluted data.
- Only extract specific opinions for iso_language_code (India and England) and to make it domain specific. Also replace unwanted special characters with space. E.g each sentence ends with comma (,), many users have habit to use comma within the sentence. Computer will read each comma as end of statement. So, pre-processing helps in making each sentences completely recognized.

Phase 2: Store in local database
This phase basically records all online opinions where each opinion sentence recognized with its table id entry, user_id, time, country and product as shown in figure 5.

	Field	Type
☐	id	bigint(20)
☐	created_at	datetime
☐	from_user	varchar(100)
☐	from_user_id	bigint(20)
☐	text	text
☐	iso_language_code	varchar(10)
☐	query	text

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. **iso_language_code:** It represents the country code to describe about the tweet.
7. **Query:** It indicates the name of the selected product.

Why we need above database fields: All fields above mentioned are mandatory to reach at the result. Only extraction of text (tweets or opinion sentences) is not the job but each detail concerns the opinion also matter. E.g. who is the user posting opinions, when opinion was posted, in which country and for which product the opinion was posted?

Phase 3: Create taxonomy of different products (databank of product keywords)
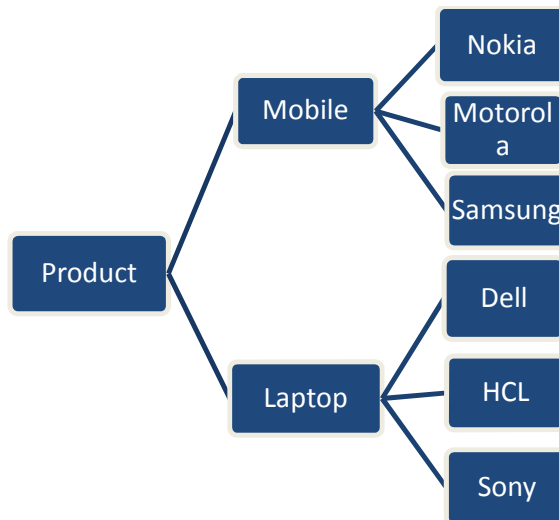

Figure 6: The taxonomy of products

Phase 4: Create databases of positive and negative opinion words
Opinions words (opinion polarity) database can be of two types.
Positive words database
Negative words database

Positive words database: A database including only positive sense words like amazing, anxious, awesome, beautiful, attractive, best, better, calm, brilliant, charm, comfort, easy, fabulous, extraordinary, fame, glowing, graceful, high, hallmark, impressive, good, wonderful etc as shown below in table 1.

Table 1: Positive opinion words database

Absolutely	Best ever	confident	Excellent	Glad
Abundant	better	Consummate	exceptional	Gladden
	Better than ever	content	Exceptional	Gladly
accepting	Bewildering	courageous	excited	Gladness
	Beyond belief	daring	Exciting	Glamorous
accomplish	Beyond compare	delighted	expressive	Glee
alive	blessed	Delighted	Extraordinary	gleeful
amaze	Blissful	Delightful	Extravagant	Gleeful
amazed	bold	determined	Extreme	Gleefully
amazement	Boundless	Determined	Fabulous	glimmering
amazes	brave	dynamic	Fabulous	Glorious
Amazing	bright	eager	Fabulous	Glow
amazingly	bright	earnest	Fabulously	Glowing
animated	bright	ease	Famed	Goodwill
anxious	brilliant	eased	Famous	Good
Astonishing	brilliant	eases	Fine	Gooood
Astounding	Brilliant	easier	Famously	Gorgeous
at ease	Brilliant	easiest	Famously	Gorgeously
attract	calm	easiness	fantastic	Grace
attractive	captivate	easing	fantastic	Graceful
Awe-inspiring	certain	easy	fascinate	Gracefully
Awesome	certain	easy	fascinating	Gracious
Beaming	challenged	easygoing	Fast	Graciously
beautiful	charm	easy-to-use	Faster	Graciousness
beautiful	charm	ecstatic	Fastest	Grand
beautifully	charming	Ecstatic	Fastest-growing	Grand
beautify	charming	elated	Fast-growing	Grandeur
beauty	charmingly	Elated	Fast-paced	Grateful
believable	chaste	impress	Faultless	Gratefully
beloved	cheaper	Impressed	festive	Gratification
benefactor	cheapest	happy	fine	great
beneficent	cheerful	Hard to believe	Fine-looking	Great
beneficial	clear	Hardy	Finely	Hallmark
beneficially	clever	Nice	Finer	Colorful
beneficiary	high	ultimate	Finest	
benefit				
benefits				

Phase 5: Create a score matrix
Score matrix or scoring matrix is a match between "the most similar elements". It is a matrix of score values which gives a weighted match to any given substring of fixed length. It has a combination of rows and columns. Rows represent opinions words from databases of positive

and negative opinion words as shown in figure &. The column values represent opinion words extracted from online opinion sentences.

Figure 7: A Score matrix located for four positive opinion words

In this research, a score matrix for extracted web opinions corresponds to four positive opinion words has been created as shown in figure 7 which gives score output for following opinion words as:

Opinion Word	Score Value
good	4
ultimate	3
happy	1
great	2

It indicates that the user who has posted this particular tweet, he/she has used good word four times, ultimate word three times, happy word one time and great word two times.

CONCLUSION

This paper provides a novel method to find score value for product opinions correspond to positive and negative opinion words. A lexical database of opinion words has been used and online opinion (twitter tweets in this research for Nokia mobile product) sentences has been extracted using API. Basically score value is created in a matrix with a score value. Increasing score value means high use of that particular opinion word in online opinions.

REFERENCES

[1] Ramandeep Sandhu et. Al. Opinion Mining using unequal Grouping for Product/Services Opinion in International Conference on "Advancements in Computing and Communication" held at Baba Banda Singh Bahadur Engineering College, Fahehgarh Sahib Punjab from February 23-25, 2012.

[2] Ramandeep Sandhu and Rahul Mehta Applying Opinion Mining to Organize Web Opinions in International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol. 1, No. 4, August 2011.

[3] Extraction and decomposition of Feature Based Web Opinions- Ramandeep Sandhu and Kiran Jyoti, National Seminar on Advances in Computer Applications April 28-29, 2011.

[4] Opinion Mining: How to Manage Opinions on Web- Ramandeep Sandhu, Harpreet Kaur and Pinderjit Kaur, National Conference on Innovative Products at Baddi University, Solan March, 26 2011.

[5] Polarity decomposition of Feature based Web Opinions- Ramandeep Sandhu and Kiran Jyoti, National Conference on Latest Advancements in Science, Engineering & Research May, 26-27 2011.

[6] Opinion Extraction and Summarization on the Web Minqing Hu and Bing Liu American Association for Artificial Intelligence (www.aaai.org).

[7] Opinion Mining & Summarization - Sentiment Analysis Bing Liu Tutorial given at WWW-2008, April 21, 2008 in Beijing.

Author

Er. Ramandeep Sandhu is pursuing M. Tech (Thesis left) in Computer Science Engineering from Guru Nanak Dev Engg College, Ludhiana with ag g. 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 4.2 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.