International Journal of Electrical and Computer Engineering (IJECE)
Vol. 5, No. 5, October 2015, pp. 1180~1187
ISSN: 2088-8708

Online Crowds Opinion-Mining it to Analyze Current Trend: A Review

Haritha Akkineni*, P.V.S. Lakshmi, B. Vijay Babu*****
* Department of Computer Science and Engineering, KL University, India
** Dept of Information Technology, PVP Siddhartha Institute of Technology, Vijayawada, India
*** Dept of CSE, KL University, India

Article Info

Article history:

Received Apr 30, 2015
Revised Jun 13, 2015
Accepted Jun 30, 2015

Keyword:

Crowd Sourcing
Micro Blogging
Opinion Mining
Social Networks
Word Sense Disambiguation

ABSTRACT

Online presence of the user has increased, there is a huge growth in the number of active users and thus the volume of data created on the online social networks is massive. Much are concentrating on the Internet Lingo. Notably most of the data on the social networking sites is made public which opens doors for companies, researchers and analyst to collect and analyze the data. We have huge volume of opinioned data available on the web we have to mine it so that we could get some interesting results out of it with could enhance the decision making process. In order to analyze the current scenario of what people are thinking focus is shifted towards opinion mining. This study presents a systematic literature review that contains a comprehensive overview of components of opinion mining, subjectivity of data, sources of opinion, the process and how does it let one analyze the current tendency of the online crowd in a particular context. Different perspectives from different authors regarding the above scenario have been presented. Research challenges and different applications that were developed with the motive opinion mining are also discussed.

Corresponding Author:

Haritha Akkineni,
Research Scholar, Dept of CSE, KLU,
74-13/1-24A, Plot no: 246, New RTC Colony,
Vijayawada, Andhra Pradesh, India 520007.
Email:akkinenih@gmail.com

1. INTRODUCTION

The increased penetration of the internet among the folks has made it a wide indispensible channel for people to communicate. There is much growth of the World Wide Web not only in its size but also in terms of services and the content that is being provided. There is a huge growth in the number of active users and thus the volume of data created on the online social networks is massive. "Twitter, a popular micro-blogging service, has 288 million monthly active users, who post about 500 million tweets a day" [1]. This itself depicts that the enormous amount of user generated content is being generated every day.

There has been a drift in how the information is being managed and shared. From only just consuming the available content to annotating it and generating new information. There can be different ways like comments on the exiting information, bookmark pages, provide ratings, share their ideas with community at large.

Online presence of the user has increased, Much are concentrating on the Internet Lingo. Notably most of the data on the social networking sites is made public which opens doors for companies, researchers and analyst to collect and analyze the data. Researchers are monitoring on the trending topics, memes and some notable events including political events [2], stock market fluctuations [3], disease epidemics [4] etc to source out the voice of crowd and observe the present tendency of online crowds towards a particular issue.

We have huge volume of opinioned data available on the web we have to mine it so that we could get some interesting results out of it with could enhance the decision making process.

Recognizing and evaluation of various viewpoints has got its practicality in various domains. Decision makers in governments and political entities need to know how the public retort to the decisions taken by them, funding agencies need to gauge their success, business need to know what people think of their products, The impact of a researchers work can be calculated by hiring committees of universities and research institutions. Such postings have also mobilized masses for political changes such as those happened in some Arab countries in 2011. It has thus become a necessity to collect and study opinions on the Web.In order to analyze the current scenario of what people are thinking focus is shifted towards opinion mining.

This study presents a systematic literature review that contains a comprehensive overview of components of opinion mining, subjectivity of data, how it relates to sentiment analysis, sources of opinion. The process and how does it let one analyze the current tendency of the online crowd in a particular context. Different perspectives from different authors regarding the above scenario has been presented. Research challenges and different applications that were developed with this motive opinion mining are also discussed.

2. LITERATURE REVIEW

2.1 Opinion Mining

It emerges from the basic field of text mining [5], a subset generally text in unstructured format to organize it in a proper way and to mine some useful information from it. Opinions are the key influencers of our behavior which supplements the decision making process [6].

Opinion: An opinion is a quintuple, (ei; aij ; ooijkl; hk; tl),
where ei is the name of an entity,
aij is an aspect of ei,
ooijkl is the orientation of the opinion about aspect aij ofentity ei,
hk is the opinion holder,
and tl is the time when the opinion is expressed by hk.

The opinion orientation ooijkl can be positive, negative or neutral, or be expressed with different strength/intensity levels.
To carry out the general process one needs to perform the following tasks:

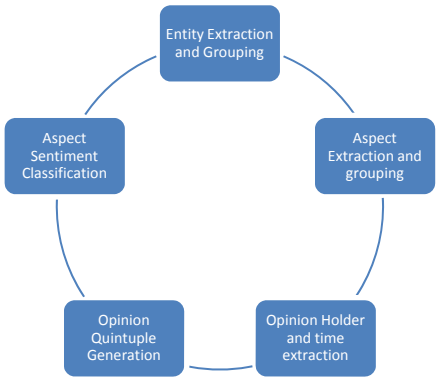

Figure1. Tasks involved

2.2 Components of Opinion Mining

Source of the opinion:

Obtaining public and consumer opinions has long been a huge trade itself for marketing, public relations, and political campaign companies. The scenario of asking others for their perspective had shifted towards online crowd sourcing [7]. There are different forms like conducting surveys, opinion polls and focus groups.

User Generated Content: Blogging, Microblogging and social networks are the most popular forms of UGC [8].

Question-Answer Databases (eg. Yahoo Answers, Ask.com)
Digital video (Youtube, Vimeo)
Blogs (Blogger, Weebly)
MicroBlogs (Tumblr, Twitter)
Podcasting (iTunes)
ReviewSites (Yelp, TripAdvisor)
Social Networking (Facebook, MySpace)
Wikis (Wikepedia)

Along with the digital content the UGC can be a combination of open source, free software and flexible licensing or related agreements to further reduce the barricades to the collective work. All these can be the source where we could gather the perspectives of various people.

2.3 Fact and Opinion

Subjectivity analysis:

The subjective sentence expresses some perceptions, beliefs or personal feelings. There are many forms of Subjective expressions like allegations, desires, beliefs, suspicions, and speculations

Objective analysis:

The objective sentence concentrates on the facts.The existing research concentrates on the factual information which is widely used in the area of Information Retrieval. Now people are moving towards what people feel and opinion hood determination. Opinion hood determination is divided into two sub tasks Subjectivity classification and opinion polarity classification which is in turn ripped into opinion and non opinion, positive and negative respectively. [9] The method they propose is using minimum cuts to produce subjective extracts from the text. The work has been focused in the sentence level subjectivity extraction. The processing overhead of large amount of data can be eliminated [10].

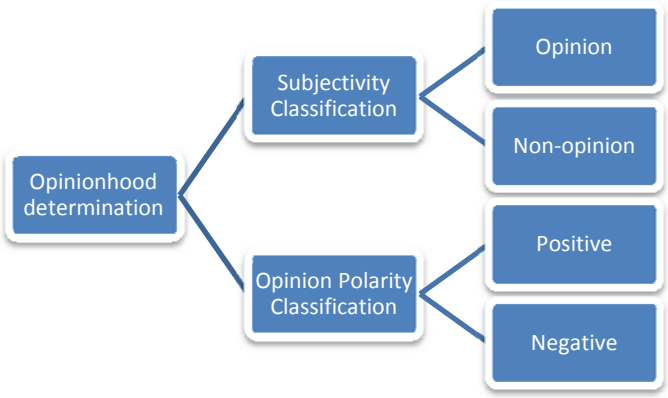

Figure 2. Opinion hood determination

2.4 Approaches Adapted

The following are the approaches adapted for opinion mining: [11]

1. Heuristics:

This approach can produce the results within a realistic timeframe. They are likely to produce the results themselves but are mostly used with the optimized algorithms.

2. Discourse Structure:

Focuses on the given text that just communicates a message, and linking it to how that message constructs a social reality or view of the world.

3. Coarse-grained Analysis:

The tasks such as retrieving the subjective documents from a collection, most recent work on this topic has focused on classification of entire document by overall positive and negative polarity.

4. Fine grained Analysis:

Tasks such as determining the attitude of a particular person on a particular topic. It is a term level analysis as it identifies whether the term is positive or negative oriented.

5. Aspect level approach:

Opinions consisted of targets and the aspects associated with them.

6. Key word analysis:

This approach classifies text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored.

7. Concept analysis:

It concentrates on semantic analysis of text through the use of web ontologies or semantic networks. The conceptual and affective information associated with natural language opinions are aggregated.

3. RESEARCH METHOD

Phase 1: Opinion extraction.
Phase 2: Processing using different techniques
Phase 3: Analyzing them to assess the current trend

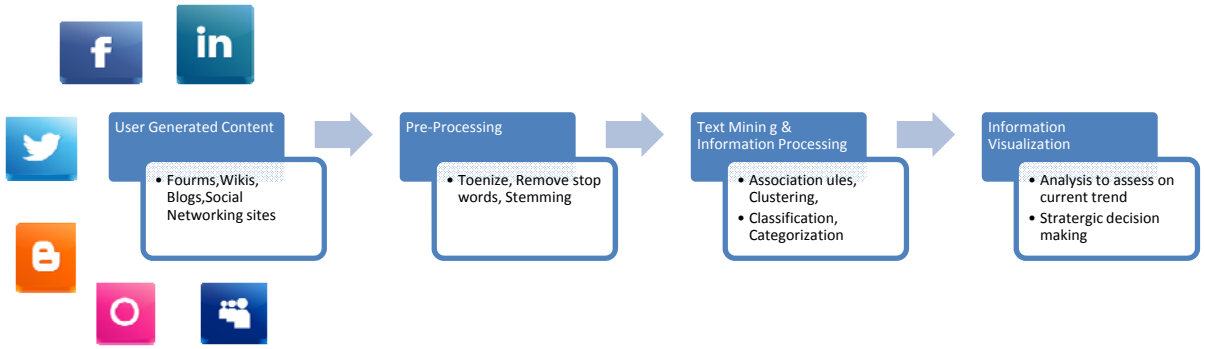

Figure 3. Schematic procedure of Opinion Mining

3.1 The list of review articles surveyed

This table contains various findings on opinion mining from different authors perspectives. It clearly states the methodology, algorithm used, data sources and the results and conclusions that were drawn.

Table 1. Articles Reviewed

Ref. No	Study	Year	Data scope	Algorithmn	Methodology used	Data Source	Conclusion	Result
[12]	Yao Wu et al	2015	Hotel review data and restaurants	Machine Learning	Combination of aspect-based opinion mining and collaborative filtering to collectively learn users' preferences on different aspects.. A unified probabilistic model Factorized Latent Aspect Model (FLAME) has been proposed.	Trip Advisor hotel review data6 and Yelp review data7	Aspect-based opinion mining techniques, aspect-based review summarization for products product level summarization-review-level analysis. Predicts the preferences based on previous preferences	FLAME outperforms other baseline models on all the measures except RMSE. The gain is especially significant compared to LRR+PMF where there are about 90% improvement on _I and 40% improvement.
[13]	Zheng Xiang et al	2015	Hotel guest experience and satisfaction	Machine Learning / statistical analysis	Classifying large amount of online customer reviews, assess the quality of these data, as well as identify inherent relationships between two domains of variables in hotel management through text analytics.	consumer reviews extracted from Expedia.com	Is study sets an example for the development of business analytics in hospitality marketing and management Or, at least, these findings can be seen as testable propositions derived from the analysis.	The underlying factors representing a set of only34 words can explain nearly 63% of the total variance in guest satisfaction, which considerably exceeded the acceptable range.

[14]	Wenting Tu,et al	2015	Shanghai market Index (SHI)	Rule based methods	Reference time (RT) information in the formation of future predictions constructed a prediction model that uses the RT information.	50,169 microblogs of investors posted on Sina Weibo.	Predictions made using reference time (RT)	Time-sensitive model, which is based on RT tags performs better than Baseline and Traditional models.
[15]	Subasish Das et al	2015	Capital Bike share of Washington DC	Lexicon	Applying sentiment analysis	Twitter	To get subjective information Text categorization with valence annotation was used.	The positive responses towards the current system were higher in frequency than the negative ones
[16]	Thuy T.B. Luong et al.	2015	Rail transit line from travelers perspective	Opinion Lexicon in English provided by Hu and Liu	Exploratory analysis and word clustering analysis was performed	Tweets from May 1st 2014 to October 1st 2014 commuter	Interactive online interface which displays and monitor real-time feedback and sentiment helpful for transit service providers.	Mondays had more positive tweets40% Tuesday more negative Tweets-55%
[17]	Yan Zhao et al	2015	Chinese hotel reviews.	Machine learning to extract aspects.	Different dimensions of features, feature representation methods and classifiers to analyze the integral effect of aspect extraction.	Online reviews	ME is the best machine learning method for aspect extraction of Chinese hotel reviews	The highest accuracy of different supervised learning method are 93.47% (ME), 87.82% (NB) and 86.34% (SVM).
[18]	Chetashri hadane et al	2015	Mobile domain	Machine learning (SVM) combined with domain specific lexicons.	Two-step method aspect classification followed by polarity classification	Online reviews	A set of techniques like linear kernel for aspect classification and polarity identification of products	Linear Kernel backed up with maximum accuracy. The accuracy achieved was 78.05% testing was done using 41 reviews.
[19]	Rahul Tejwani et al	2014	Academic	Lexicon based features	Thayes Model of human emotion	Yelp Review dataset	Classify text in two dimensional space on polarity and intensity	Polarity: T mean accuracy 81.60% +/- 1.92% intensity: mean accuracy 67.14% +/- 1.22%.
[20]	Nathan Aston et al	2014	Sanders Corpus STS-Gold and Senti Strength	Machine Learning	Modified Balanced Winnow to train on pre-labeled instances using a streaming algorithm and processing them in real time.	public datasets Twitter,My Space, Youtube,B BC, Digg	Removed all but the top features when performing classification.	Achieved the highest accuracy of 87.5% with STS_Gold on 5 grams representation.

3.2 Applications Developed

1. IFeel:

It is a web application that that detects sentiments in any form of text including unstructured social media data. A combined Method was developed. Using this tool the user can analyse the given data with combined method as well as the seven existing sentiment analysis methods like: SentiWordNet, Emoticons, PANAS-t, SASA, Happiness Index, Sentic-Net, and SentiStrength. [21]

2. Social Mention:

It is a social media search and analysis platform that aggregates user generated content from across the universe into a single stream of information. It helps us in tracking and measuring what people say. [22]

3. Sentiment140

Sentiment 140 is used to discover the sentiment of a brand, product, or topic on Twitter. Classifiers are build using machine learning algorithms. [23]

4. Opinion Finder:

It identifies subjective sentences and marks various aspects of subjectivity in these sentences, including the source of the subjectivity and words that are included in phrases expressing positive or negative sentiments. [24]

5. Trackur

It is a social media monitoring service which enables us to generate specific searches to run across a number of social media platforms. Tracks posts from Twitter, Facebook, Reddit, Delicious, Google+,blogs, videos including YouTube, and photos including Flickr [25].

Table 2. Tools used for a particular method

Method	Tools Used
Search	Google, Topsy, Social Mention, Twitter, FB
Monitoring	Google Alert, Trackur, Attentio, Engagor, Viral Heat, Radian6, Hub Spot
Track Performance	Google Analytics, Post Rank, Strutta,Your Profile, Swix, Klout, Peerindex, Grader

Search solutions are provided by Google, for following news, blog posts, videos and images the Social Mention, Addict-o-matic etc play a prominent role.Technorati.com facilitates in searching blogs, and BotBox, for online news content, including blogs .

Analytical tools for Twitter are Tweet Archivist, 140kit and TweetDeck, is an aggregation tool. Twapperkeeper, is an API for tracking Twitter activities.

An open-source command-line tool to further process data-Gawk e.g. tweet statistics and metrics, and the open source visualization tools Wordle, used also for YouTube video tags, and Gephi for visualizing networks, used for e.g. Twitter datasets and blog. The expansion of tools concerning social media examining is swift and there is a ample scope for many new offsprings to arise. [26]

3.3 Research Challenges

Coreference resolution:

The process of finding all expressions that refer to the same entity in a text It is important challenge for opinion mining as without considering a great deal of opinion information will be lost, and opinions may be assigned to wrong entities. [26]

Negation handling:

The grammatical rules used in the textual data often contain negations used in text that completely change the meanings of words. Detecting its scope with in a sentence (text) are necessary in finding out the sentiments from a piece of text. [27]

Word sense disambiguation:

Most valuable concept to be addressed with respect to sentiment analysis. The non literal senses, such as metaphors and expanded senses, tend to indicate subjectivity, triggering polarity and effecting the resulted opinion summarization.[28]

Domain adaptation problems:

The sentiment classifier trained with the labeled data from one domain normally cannot perform up to the mark in another domain. This problem is termed as the domain adaptation problem in sentiment classification by using some labeled data from the source domain and a large amount of unlabeled data from the target domain. [29]

4. CONCLUSION

This work presents an in-depth background study about opinion mining. The subject has fascinated considerable concentration since the 1990s, in particular with respect to subjectivity analysis and lexical resource generation. Based on the survey done we have seen that in 2013 focus was given to development of model and frameworks for sentiment analysis and in 2014 it has shifted to content extraction and classification and the current year focus is mostly heading towards aspect based prediction, which could be much useful in the semantic web and common sense knowledge. A number of computational models and linguistic features related to opinion mining, component analysis and opinion-target identification are thoroughly discussed.

REFERENCES

[1] Jack Dorsey, Twitter Usage, 2015, Available at https://about.twitter.com/company.

[2] A. Tumasjan, T.O. Sprenger, P.G Sandner, and I.M. Welpe. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In International AAAI conference on Weblogs and Social Media (ICWSM), 2010.
[3] J. Bollen, H. Mao, and X.J. Zeng. Twitter Mood Predicts the Stock Market. CoRR, abs/1010.3003, 2010.
[4] A. Lamb, M.J. Paul, and M. Dredze. Separating Fact from Fear: Tracking Flu Infections on Twitter. In Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 789,795, June 2013.
[5] Ionia Veritawati, Ito Wasito, T Basaruddin,Text Preprocessing using Annotated Suffix Tree with Matching Keyphrase, Vol 5, No 3, International Journal of Electrical and Computer Engineering, 2015.
[6] J. Trout and S. Rivkin, The New Positioning: The Latest on the World's #1Business Strategy. New York: McGraw Hill, 1996.
[7] Jery Althaf, Sreevas S, Crowdsourcing Platform For Approching Humanitarian Challenges, Internationl Journal of Social Networking and Virtual Communities,Vol1, No2, Dec 2012,
[8] Boundless. "Types of Consumer Generated Digital Content" Boundless MArketing, 3, Jul, 2014, retrieved from https://www.boundless.com/marketing/textbooks/boundless-marketing-textbook/
[9] Ge Xu, Houfeng Wang, 2011. Semi-supervised Chinese contextual polarity classification with automatic feature selection. Paper presented at the 2011 Seventh International Conference on Natural Computation (ICNC).
[10] B. Pang and L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts.
[11] Cambria, E., Hussain, A., 2012. Sentic computing: Techniques, Tools, and Applications, Springer Netherlands, ISBN: 978-94-007-5069-2.
[12] Yao Wu, Martin Ester, FLAME: A Probabilistic Model Combining Aspect Based Opinion Mining and Collaborative Filtering, WSDM'15, February 2–6, 2015, Shanghai, China. Copyright 2015 ACM 978-1-4503-3317-7/15/02 http://dx.doi.org/10.1145/2684822.2685291.
[13] Zheng X, Zvi S, John H.G, Muzaffer U, What can big data and text analytics tell us about hotel guestexperience and satisfaction?, International Journal of Hospitality Management 44 (2015) 120–130, http://dx.doi.org/10.1016/j.ijhm.2014.10.013 0278-4319/© 2014 Elsevier Ltd.
[14] Wenting Tu, David Cheung, and Nikos Mamoulis, Time-sensitive Opinion Mining for Prediction, 2015, Association for the Advancement of Artificial Intelligence
[15] Das, S., Sun, X.D. and Dutta, A. (2015). Investigating User Ridership Sentiments for Bike Sharing Programs. Journal of Transportation Technologies, 5, 69-75. http://dx.doi.org/10.4236/jtts.2015.52007
[16] Thuy T.B. Luong, Douglas Houston, Public opinions of light rail service in Los Angeles, an analysis using Twitter data, iConference 2015 Proceedings
[17] Yan Zhao1, Suyu Dong2 and Jing Yang, Effect Research of Aspects Extraction for Chinese Hotel Reviews Based on Machine Learning Method, International Journal of Smart Home Vol. 9, No. 3 (2015), pp. 23-34 http://dx.doi.org/10.14257/ijsh.2015.9.3.03
[18] Chetashri Bhadanea, Hardi Dalalb, Heenal Doshi, Sentiment analysis: Measuring opinions, International Conference on Advanced Computing Technologies and Applications (ICACTA-2015), proceedings.
[19] Rahul Tejwani, Twodimensional Sentiment Analysis of text, arXiv:1406.2022v1 [cs.IR] 8 Jun 2014
[20] Nathan A, Timothy M, Jacob L, Garrett H, Dane L,Wei Hu, Sentiment Analysis on the Social Networks Using Stream Algorithms, Journal of Data Analysis and Information Processing, 2014, 2, 60-66,Published Online May 2014 in SciRes http://www.scirp.org/journal/jdaip,http://dx.doi.org/10.4236/jdaip.2014.22008
[21] Matheus A, Pollyanna G, Meeyoung C, Fabrício B, iFeel: A Web System that Compares and Combines Sentiment Analysis Methods, International World Wide Web Conference Committee (IW3C2). April 7–11, 2014, Seoul, Korea. ACM 978-1-4503-2744-2/14/04.
[22] Linus Philip Lawrence, Reliability of Sentiment Mining Tools: A comparison of Semantria and Social Mention, 3rd IBA Bachelor Thesis Conference, July 3rd, 2014, Enschede, The Netherlands.
[23] Felipe B, Marcelo M, Barbara P, Combining Strengths, Emotions and Polarities for Boosting Twitter Sentiment Analysis, WISDOM'13, August 11 2013, Chicago, IL, USA Copyright 2013 ACM 978-1-4503-2332-1/13/08.
[24] Ben He, Craig Macdonald, Iadh Ounis, Ranking Opinionated Blog Posts using OpinionFinder Copyright is held by the author/owner(s). SIGIR'08, July 20–24, 2008, Singapore. ACM 978-1-60558-164-4/08/07.
[25] Aino R, MaritaV, Social Media Monitoring for Crisis Communication: Process, Methods and Trends in the Scientific Literature, Online Journal of Communication and Media Technologies Volume: 4 – Issue: 1 – January - 2014
[26] Ding, X. and B. Liu. Resolving object and attribute coreference in opinion mining. In Proceedings of International Conference on Computational Linguistics (COLING-2010), 2010.
[27] Councill, I.G., Mcdonald, R. and Velikovich, L. (2010). What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In: The Workshop on Negation and Speculation in Natural Language Processing. Uppsala, Sweden.Association for Computational Linguistics, 51-59.
[28] V. Rentoumi, George G. Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach. Proccedings of International Conference RANLP 2009 - Borovets, Bulgaria, pages 370–375, 2009.
[29] Shoushan Li, Yunxia Xue, Zhongqing Wang, Guodong Zhou, Active Learning for Cross-omain Sentiment Classification, Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

BIOGRAPHIES OF AUTHORS

Mrs. Haritha Akkineni received her Master of Information Technology degree from University of Ballarat in 2005.She is pursuing Ph.D (CSE) from Koneru Lakshmaiah University. Her interested areas are Data Mining, Image Mining, Opnion Mining, Sentiment analysis and Social Networks. Published papers in various reputed International Journals including Springer.

Dr. B.Vijaya Babu, B.Tech, M.Tech, Ph.D is working as a professor at KL University, He has got his Ph.D from Andhra University in year 2012, He has got 20 years of teaching experience. His area of research is Data Mining/Knowledge Engineering and published more than 25 research publications in various National and International Journals, including SCOPUS indexed.

Dr P.V.S Lakshmi, B.Tech, MS, Ph.D working as professor in PVP Siddhartha Institute of Technology Vijayawada. She has got 21 years of teaching experience and 6 years of research experience. She was awarded with Ph.D in the year 2011in the area of Arificial Intelligence and Image Processing. She has published papers in various reputed journals including Springer.Her areas of interests are Data Mining, Database Management Systems, Image Processing, Social Networks and Sentiment Analysis.