



POSITION SUPPOSITION FOR NON ENVIRONMENTAL TAGGED TWEETS IN USER TIMELINES

Prof.K.P.Ragavi,^[1]A.Babu^[2], V.Sangametha^[3], R.Soundharya^[4],
Assistant Professor^[1], UG Scholar ^[2]^[3]^[4]
Department of Information Technology,

Sengunthar College of Engineering, Tiruchengode, Tamil Nadu, India

^[1]ragaviprakashmail1@gmail.com, ^[2]krishrai266@gmail.com, ^[3]sangametha123@gmail.com,
^[4]soundharyarajan96@gmail.com

Manuscript History

Number: **IJIRAE/RS/Vol.07/Issue03/Special Issue/33.MRITSCE10112**

Received: 15, February 2020

Final Correction: 27, February 2020

Final Accepted: 10, March 2020

Published: **14, March 2020**

Editor: Dr.A.Arul Lawrence selvakumar, Chief Editor, IJIRAE, AM Publications, India

Copyright: ©2020 This is an open access article distributed under the terms of the Creative Commons Attribution License, Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Abstract - We propose a collaborative multi-Trends sentiment classification approach to train sentiment classifiers for multiple tweets simultaneously specifically; we decompose the sentiment classifier of each Trend into two components, a global one and a Trends-specific one. Our method provides an efficient way to accurately categorize trending topics without need of external data, enabling news organizations to discover breaking news in real-time, or to quickly identify viral memes that might enrich marketing decisions, among others. The global model can capture the general sentiment knowledge and is shared by various tweets. The Trends-specific Greedy & Dynamic Blocking Algorithms model can capture the specific sentiment expressions in each trend. Experimental results on benchmark datasets show that our approach can effectively improve the performance of multi-Trends sentiment classification and significantly outperform baseline methods.

Keywords: global model, Dynamic Blocking, Trends-specific Greedy

INTRODUCTION

Web Opinion Data Mining Concept

Web Review Analysis and Opinion Mining

The development of Web 2.0 websites, user generated content (UGC), such as product reviews, blogs, microblogs and so on, has been growing explosively. Mining the sentiment information in the massive user generated content can help sense the public's opinions towards various topics, such as topics, brands, disasters, events, celebrities and so on, and is useful in many applications. For example, researchers have found that analyzing the sentiments in tweets has the potential to predict variation of stock market prices and presidential election results. Classifying the sentiments of massive microblog messages is also helpful to substitute or supplement traditional polling, which is expensive and time-consuming. Product review sentiment analysis can help companies improve their topics and services, and help customers make more informed decisions. Analyzing the sentiments of user generated content is also proven useful for user interest mining, personalized recommendation, social advertising, customer relation management, and crisis management. Thus, sentiment classification is a hot research topic in both industrial and academic fields.

OVERVIEW OF DATA MINING

Data mining (sometimes called data or knowledge discovery) is the progression of analyses data from special perspectives and abbreviation into useful data information that can be used enlarge the revenue, reduce cost and both. Data mining software is individual a number of logical tools for analyzing the data's. It allows the users to analyzed data from various dimensions or angles and review the associations recognized.

Technically, the data mining is the process of decision correlations or patterns between fields in huge relational databases. In this paper present a beginning to the essential technologies of data mining. Today's business environment is applicable profitable applications as well as a beginning narrative of how data warehouse architectures to develop to distribute the rate of data mining to end users.

LITERATURE REVIEW

Cross-Trends Sentiment Classification via Spectral Feature Alignment

In this paper [16] Sinno Jialin Pan, has proposed Sentiment classification aims to automatically predict sentiment polarity (e.g., positive or negative) of users publishing sentiment data (e.g., reviews, blogs). Although traditional classification algorithms can be used to train sentiment classifiers from manually labeled text data, the labeling work can be time-consuming and expensive. Meanwhile, users often use some different words when they express sentiment in different tweets. If we directly apply a classifier trained in one Trend to other tweets, the performance will be very low due to the differences between these tweets. In this work, we develop a general solution to sentiment classification when we do not have any labels in a target Trends but have some labeled data in a different Trends, regarded as source Trends. In this cross-Trends sentiment classification setting, to bridge the gap between the tweets, we propose a spectral feature alignment (SFA) algorithm to align Trends-specific words from different tweets into unified clusters, with the help of Trends independent words as a bridge. In this way, the clusters can be used to reduce the gap between Trends-specific words of the two tweets, which can be used to train sentiment classifiers in the target Trends accurately. Compared to previous approaches, SFA can discover a robust representation for cross-Trends data by fully exploiting the relationship between the Trends-specific and Trends independent words via simultaneously co-clustering them in a common latent space. We perform extensive experiments on two real world datasets, and demonstrate that SFA significantly outperforms previous approaches to cross-Trends sentiment classification.

Automatically Extracting Polarity-Bearing Topics for Cross-Trends Sentiment Classification

In this paper work [19] Yulan He, has proposed Joint sentiment-topic (JST) model was previously proposed to detect sentiment and topic simultaneously from text. The only supervision required by JST model learning is Trends-independent polarity word priors. In this paper, we modify the JST model by incorporating word polarity priors through modifying the topic-word Dirichlet priors. We study the polarity-bearing topics extracted by JST and show that by augmenting the original feature space with polarity-bearing topics, the in-Trends supervised classifiers learned from augmented feature representation achieve the state-of-the-art performance of 95% on the movie review data and an average of 90% on the multi-Trends sentiment dataset. Furthermore, using feature augmentation and selection according to the information gain criteria for cross-Trends sentiment classification, our proposed approach performs either better or comparably compared to previous approaches. Nevertheless, our approach is much simpler and does not require difficult parameter tuning.

Regularized Multi-Task Learning

In this paper work [20] Theodoros Evgeniou, has proposed Past empirical work has shown that learning multiple related tasks from data simultaneously can be advantageous in terms of predictive performance relative to learning these tasks independently. In this paper we present an approach to multi-task learning based on the minimization of regularization functional similar to existing ones, such as the one for Support Vector Machines (SVMs), that has been successfully used in the past for single-task learning. Our approach allows modeling the relation between tasks in terms of a novel kernel function that uses a task-coupling parameter. We implement an instance of the proposed approach similar to SVMs and test it empirically using simulated as well as real data. The experimental results show that the proposed method performs better than existing multi-task learning methods and largely outperforms single-task learning using SVMs.

Opinion mining and sentiment analysis

In this paper work [1] Bo Pang, has proposed an important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object. This survey covers techniques and approaches that promise to directly enable opinion-oriented information seeking systems.



Our focus is on methods that seek to address the new challenges raised by sentiment aware applications, as compared to those that are already present in more traditional fact-based analysis. We include material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena

In this paper work [2] Johan Bollen has proposed we perform a sentiment analysis of all tweets published on the microblogging platform Twitter in the second half of 2008. We use a psychometric instrument to extract six mood states (tension, depression, anger, vigor, fatigue, confusion) from the aggregated Twitter content and compute a six-dimensional mood vector for each day in the timeline. We compare our results to a record of popular events gathered from media and sources. We find that events in the social, political, cultural and economic sphere do have a significant, immediate and highly specific effect on the various dimensions of public mood. We speculate that large scale analyses of mood can provide a solid platform to model collective emotive trends in terms of their predictive value with regards to existing social as well as economic indicators.

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

In this paper work [3] Brendan O'Connor, has proposed We connect measures of public opinion measured from polls with sentiment measured from text. We analyze several surveys on consumer confidence and political opinion over the 2008 to 2009 period, and find they correlate to sentiment word frequencies in contemporaneous Twitter messages. If we want to know, say, the extent to which the U.S. population likes or dislikes Barack Obama, an obvious thing to do is to ask a random sample of people (i.e., poll). Survey and polling methodology, extensively developed through the 20th century (Krosnick, Judd, and Wittenbrink), gives numerous tools and techniques to accomplish representative public opinion measurement. In this paper, we connect measures of public opinion derived from polls with sentiment measured from analysis of text from the popular microblogging site Twitter. We explicitly link measurement of textual sentiment in microblog messages through time, comparing to contemporaneous polling data. In this preliminary work, summary statistics derived from extremely simple text analysis techniques are demonstrated to correlate with polling data on consumer confidence and political opinion, and can also predict future movements in the polls.

Mining and summarizing customer reviews

In this paper work [4] Mingqing Hu, has proposed Merchants selling topics on the Web often ask their customers to review the topics that they have purchased and the associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product. It also makes it difficult for the manufacturer of the product to keep track and to manage customer opinions. For the manufacturer, there are additional difficulties because many merchant sites may sell the same product and the manufacturer normally produces many kinds of topics. In this research, we aim to mine and to summarize all the customer reviews of a product.

Opinion Flow: Visual Analysis of Opinion Diffusion on Social Media

In this paper work [6] Yingcai Wu, has proposed it is important for many different applications such as government and business intelligence to analyze and explore the diffusion of public opinions on social media. However, the rapid propagation and great diversity of public opinions on social media pose great challenges to effective analysis of opinion diffusion. In this paper, we introduce a visual analysis system called Opinion Flow to empower analysts to detect opinion propagation patterns and glean insights. Inspired by the information diffusion model and the theory of selective exposure, we develop an opinion diffusion model to approximate opinion propagation among Twitter users. Accordingly, we design an opinion flow visualization that combines a Sankey graph with a tailored density map in one view to visually convey diffusion of opinions among many users. A stacked tree is used to allow analysts to select topics of interest at different levels. The stacked tree is synchronized with the opinion flow visualization to help users examine and compare diffusion patterns across topics. Experiments and case studies on Twitter data demonstrate the effectiveness and usability of Opinion Flow.

Twitter Sentiment Classification using Distant Supervision

In this paper work [8] Alec Go, has proposed we introduce a novel approach for automatically classifying the sentiment of Twitter messages. These messages are classified as either positive or negative with respect to a query term.

This is useful for consumers who want to research the sentiment of topics before purchase, or companies that want to monitor the public sentiment of their brands. There is no previous research on classifying sentiment of messages on microblogging services like Twitter. We present the results of machine learning algorithms for classifying the sentiment of Twitter messages using distant supervision. Our training data consists of Twitter messages with emoticons, which are used as noisy labels. This type of training data is abundantly available and can be obtained through automated means. We show that machine learning algorithms (Greedy & Dynamic Blocking Algorithms, Maximum Entropy, and SVM) have accuracy above 80% when trained with emoticon data. This paper also describes the preprocessing steps needed in order to achieve high accuracy. The main contribution of this paper is the idea of using tweets with emoticons for distant supervised learning.

A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems

In this paper work [14] Amir Beck, has proposed we consider the class of iterative shrinkage-thresholding algorithms (ISTA) for solving linear inverse problems arising in signal/image processing. This class of methods, which can be viewed as an extension of the classical gradient algorithm, is attractive due to its simplicity and thus is adequate for solving large-scale problems even with dense matrix data. However, such methods are also known to converge quite slowly. In this paper we present a new fast iterative shrinkage-thresholding algorithm (FISTA) which preserves the computational simplicity of ISTA but with a global rate of convergence which is proven to be significantly better, both theoretically and practically. Initial promising numerical results for wavelet-based image deblurring demonstrate the capabilities of FISTA which is shown to be faster than ISTA by several orders of magnitude. Linear inverse problems arise in a wide range of applications such as astrophysics, signal and image processing, statistical inference, and optics, to name just a few. The interdisciplinary nature of inverse problems is evident through a vast literature which includes a large body of mathematical and algorithmic developments; see, for instance, the monograph and the references therein. A basic linear inverse problem leads us to study a discrete linear system of the form

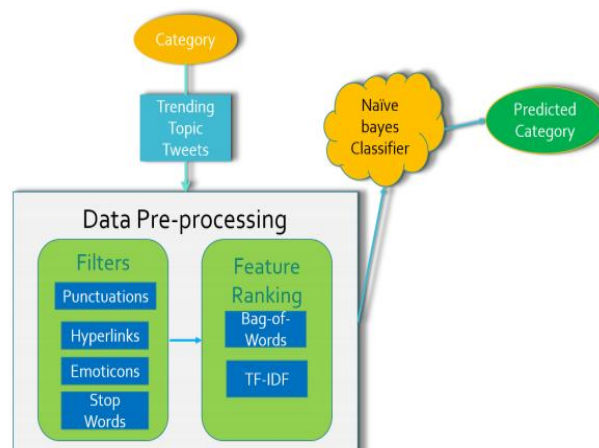
$$Ax = b + w$$

Where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are known, w is an unknown noise (or perturbation) vector, and x is the “true” and unknown signal/image to be estimated. In image blurring problems, for example, $b \in \mathbb{R}^m$ represents the blurred image, and $x \in \mathbb{R}^n$ is the unknown true image, whose size is assumed to be the same as that of b (that is, $m = n$). Both b and x are formed by stacking the columns of their corresponding two-dimensional images. In these applications, the matrix A describes the blur operator, which in the case of spatially invariant blurs represents a two-dimensional convolution operator. The problem of estimating x from the observed blurred and noisy image b is called an image deblurring problem.

PROPOSED SYSTEM

In our proposed work Greedy & Dynamic Blocking Algorithms recommends tweets by matching users with other users having similar interests. It collects user feedback in the form of ratings provided by user for specific tweets and finds match in rating behaviors among users in order to find group of users having similar preferences. One of the main features on the homepage of Twitter shows a list of top terms so-called trending topics at all times. These terms reflect the topics that are being discussed most at the very moment on the site’s fast-flowing stream of tweets. In order to avoid topics that are popular regularly (e.g., good morning or good night on certain times of the day), Twitter focuses on topics that are being discussed much more than usual, i.e., topics that recently suffered an increase of use, so that it trended for some reason. Here, a user profile represents user preferences that the user has either explicitly or implicitly provided.

Architecture Diagram



SYSTEM DESIGN

Input Design

Input design is the process of converting the user-originated input to a computer based format. The design decision for handling input specify how data are accepted for computer processing. Input design is a part of overall system design that needs careful attention. The collection of input data is considered to be the most expensive part of the system design. Since the inputs have to be planned in such a way so as to get the relevant information, extreme care is taken to obtain the pertinent information. If the data going into the system is incorrect then the processing and outputs will magnify these errors. The goal of designing input data is to make data entry as easy, logical and free from errors as possible. The nature of input data is determined partially during logical system design. However the nature of inputs is made more explicit during the physical design. The impact of inputs on the system is also determined. Effort has been made to ensure that input data remains accurate from the stage at which it is recorded and documented to the stage at which it is accepted by the computer. Validation procedures are also present to detect errors in data input, which is beyond control procedures. Validation procedures are designed to check each record, data item or field against certain criteria. To address this problem, we classify Twitter Trending Topics into 18 general categories such as sports, politics, technology, etc. We experiment with 2 approaches for topic classification; (i) the well-known Bag-of-Words approach for text classification and (ii) network-based classification. In text-based classification method, we construct word vectors with trending topic definition and tweets, and the commonly used tf-idf weights are used to classify the topics using a Naive Bayes Multinomial classifier. In network-based classification method, we identify top 5 similar topics for a given topic based on the number of common influential users. The categories of the similar topics and the number of common influential users between the given topic and its similar topics are used to classify the given topic using a C5.0 decision tree learner. Experiments on a database of randomly selected 768 trending topics (over 18 classes) show that classification accuracy of up to 65% and 70% can be achieved using text-based and network-based classification modeling respectively. Keywords-Social Network

Output Design

The output is designed in such a way that it is attractive, convenient and informative. Forms are designed in JAVA with various features, which make the console output more pleasing. As the outputs are the most important sources of information to the users, better design should improve the system's relationships with us and also will help in decision-making. Form design elaborates the way output is presented and the layout available for capturing information. In this project, the output is designed in the form reports. The system is designed to generate various user friendly reports to help the business process. Following are the different reports supported:

CONCLUSION

In the last few decades, twitter asynchronous systems have been used, among the many available solutions, in order to mitigate information and cognitive overload problem by suggesting related and relevant tweets to the users. In this regards, numerous advances have been made to get a high-quality and fine-tuned twitter asynchronous system. Nevertheless, designers face several prominent issues and challenges. In this work, we have touched variety of topics like natural Language Processing, Text Classification, Feature selection, Feature ranking, etc. Each one of these topics was used to leverage the massive information flowing through twitter. Understanding twitter was as important as knowing the topics in question. The results of the previous experiments, led us to the conclusion that feature selection is an absolutely necessity in a text classification system. This was proved when we compared our results with a system that uses the exact same dataset without feature selection. We were able to achieve 33.14% and 28.67% improvement with bag-of-words and TF-IDF scoring techniques correspondingly. We also mentioned recognition and some opportunities that our work provides in the fields of news media, marketing and businesses in general. We hope that our work can provide a good foundation to the future of text classification in social media and to the opportunities that comes with it.

REFERENCES

1. B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1–135, 2008.
2. J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in Proc. Int. AAAI Conf. Weblogs Social Media, 2011, pp. 17–21.
3. B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series," in Proc. Int. AAAI Conf. Weblogs Social Media, 2010, pp. 122–129.
4. M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2004, pp. 168–177.
5. T. Chen, R. Xu, Y. He, Y. Xia, and X. Wang, "Learning user and product distributed representations using a sequence model for sentiment analysis," IEEE Comput. Intell. Mag., vol. 11, no. 3, pp. 34–44, Aug. 2016.



6. Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu, “OpinionFlow: Visual analysis of opinion diffusion on social media,” *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1763–1772, Dec. 2014.
7. B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: Sentiment classification using machine learning techniques,” in *Proc. ACL Conf. Empirical Methods Natural Language Process.*, 2002, pp. 79–86.
8. A.Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *Stanford Univ., Stanford, CA, USA, Project Rep. CS224N*, pp. 1–12, 2009.
9. F. Wu, Y. Song, and Y. Huang, “Microblog sentiment classification with contextual knowledge regularization,” in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2332–2338.
10. J. Blitzer, M. Dredze, and F. Pereira, “Biographies, bollywood, boom-boxes and blenders: Trends adaptation for sentiment classification,” in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, vol. 7, pp. 440–447.
11. X. Glorot, A. Bordes, and Y. Bengio, “Trends adaptation for large-scale sentiment classification: A deep learning approach,” in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 513–520.
12. S.-S. Li, C.-R. Huang, and C.-Q. Zong, “Multi-Trends sentiment classification with classifier combination,” *J. Comput. Sci. Technol.*, vol. 26, no. 1, pp. 25–33, 2011.
13. L. Li, X. Jin, S. J. Pan, and J.-T. Sun, “Multi-Trends active learning for text classification,” in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 1086–1094.
14. A.Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
15. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
16. S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, “Cross-Trends sentiment classification via spectral feature alignment,” in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 751–760.
17. Y. Lu, M. Castellanos, U. Dayal, and C. Zhai, “Automatic construction of a context-aware sentiment lexicon: An optimization approach,” in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 347356.
18. S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
20. Y. He, C. Lin, and H. Alani, “Automatically extracting polaritybearing topics for cross- Trends sentiment classification,” in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics*, 2011, pp. 123–131.
21. T. Evgeniou and M. Pontil, “Regularized multi-task learning,” in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 109–117.
22. Raja, G P& Mangai, S 2018, ‘Investigation On Optimization, Prioritizing and Weight Allocation Techniques for Load Balancing and Controlling Multimedia Traffic in Wireless Mesh Network’, *International Journal of Business Information Systems*, SCOPUS Indexed Journal (Inderscience) - (P ISSN No: 1746-0972). Published Online: 10th Feb 2020, DOI: 10.1504/IJBIS.2020.105161.IF: 0.72.
23. Raja, G P& Mangai, S 2017, ‘Firefly Load Balancing Based Energy Optimized Routing for Multimedia Data Delivery in Wireless Mesh Network’, *Cluster Computing-The Journal of Networks Software Tools and Applications*, SCOPUS Indexed Journal (Springer) - (E ISSN No: 1573-7543).Published Online: 27th Dec 2017, <https://doi.org/10.1007/s10586-017-1557-1>.IF: 2.040.
24. Geetha. E & Nagarajan. C , 2018, ‘Induction Motor Fault Detection and classification Using Current Signature Analysis Technique’ *Emerging Devices and Smart Systems*, (IEEE Xplore) NSPEC Accession Number:18302693,DOI:10.1109/ICEDSS.2018.8544272, pp:48 – 52
25. Geetha. E & Nagarajan. C , 2015, ‘Embedded controller integrated into a wireless sensor network for induction motor parameter monitoring using ZigBee, *international Journal of Applied Engineering Research*, Special Issue, Volume 10, No.9 2015, pp.9343-9346.
26. Satish Kumar.R. and Sanavullah, M.Y.,” Theoretical and Experimental identification of cooking spot for shell eggs without explosions in a domestic Microwave Oven”, *Canadian Journal on Electrical and Electronics Engineering*, Vol. 1, No. 4, pp.71-78, June 2010.
27. Satish Kumar.R, K.Uma Devi, and Sanavullah, M.Y.,” Performance Analysis of using Exterior Rotor Permanent Magnet Brushless DC (ERPMBLDC)Motor”, *Improvement by a Novel Peak Torque Excitation Technique”, International journal of Innovative research in Advanced Engineering*, Vol. 1, 2012, pp. 1-7 (Impact factor 1.311).
28. Dr.R.Satish Kumar and Dr.K.Umadevi ” Finite Element Analysis of an Exterior Rotor Permanent Magnet Brushless DC Motor for Torque Improvement by a Novel Peak Torque Excitation Technique”, *International journal of Innovative research in Advanced Engineering*, Vol. 1, 2014, pp. 1-6 (Impact factor 1.311).
29. Dr.R.Satish Kumar and Dr.K.Umadevi ” A Novel peak torque Excitation Technique for Torque Improvement in Exterior Rotor Permanent Magnet Brushless DC Motor”, *International journal of Innovative research in Advanced Engineering*, Vol. 1, 2014, pp. 227-236 (Impact factor 1.311).