iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** IX **Month of publication:** September 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 4 Issue IX, September 2016
ISSN: 2321-9653

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Online News Article Temporal Phrase Extraction for Causal Linking

Amol S Patwardhan

Senior Researcher, VIT, University of Mumbai, 400037, India.

Abstract— This research proposes a method to detect temporal links between news articles using phrase extraction and position analysis. The words from the information feed were split into tokens using special character as the delimiters. The words, symbols and tokens were marked by the type of word such as noun, verb, adjective, preposition and adverb. The position of each word and its length and distance from the verb on the either side of previous 2 and next 2 sentences was extracted as a feature. The time based relations between several sentences using noun as the root and verbs as the action event was stored. The link length and hop count was measured and the connected tree was stored as a decision model. The prediction model was evaluated using time based questions and event series queries. The results showed 67% accuracy in terms of correct event sequencing and recall rate. In the next step emotion based classification of the news article event actions was performed to test whether the weighted graph linking based on phrase position properties (length, count, location) method could be used for sentiment/emotion analysis with fear and happiness as the class labels. This evaluation showed clear thresholds in case of specific event-event and event-action pairs and their temporal relation type.

Keywords— Tokenization, linked graph, phrase position, anchor action verb

I. INTRODUCTION

In the last decade, millions of websites have emerged on the internet and many newspapers now have a dedicated website for current news feeds. There is a need to identify the temporal relation between various events. Such temporal analysis has applications ranging from historical data evaluation, forensic studies and archaeological examination. Many times the actual time specific word, data and information such as seconds, minutes, hours, days, week, month and year, is not found in the text. The system has to rely on presence of event specific words such as before, after, soon, later, at, since then, ago etc. It is important to develop ability to automatically answer questions about event occurrence using machines.

This paper focuses on answering the event chronology related questions from a news feed using decision tree model and weighted graphical links. We formulated the problem of event sequencing as given an event A the system must be able to create a chronological order of events and place the event in the correct location on the time scale. Additionally, the system must be able to identify unique events and similar events even when the token representing the event is different or separated by more than one sentence. Tatu and Srikanth [1] examined the temporal links between various events and methods to provide reasons for answering 'why' based queries and for information retrieval.

Various methods have been explored [2], [3], [4] to identify the temporal events from text with high accuracy. But identifying the temporal links between these events is a time consuming and laborious task even for human annotators. Automatically detecting the temporal sequence of events is challenging and only few studies have succeeded in accurately establishing the chronology of events. The problem with accurately detecting temporal relation between series of sentences is that the same sentence could have a different meaning. In addition, many times the context is important to evaluate the association of events with specific time and the context information is not available. Thus, an exhaustive semantic analysis needs to be performed on the narrative, text, news, blogs and articles to be able to detect and associate the correct time sequence of various events in the text. Depending on the techniques applied, whether it is computational, probabilistic, or supervised learning, the accuracy changes for time based analysis of events. Researchers [5], [6], [7], [8] have developed temporal corpus and evaluated techniques for temporal relation identification between consecutive sentence but accuracy of such systems is limited to the corpus because of lack of machine understanding about common knowledge that humans have. The studies have concentration mostly on consecutive sentence event tagging.

This paper proposed a graph based tree-linking technique with position of temporal phrases serving as weights for the links. Thus the system is able to traverse the tree given two events and estimate whether the event 1 is before, after or at the same time as event 2. The traversal also used dynamic adjustment to weights based on over lapping events to generate optimized graphs for temporal

Volume 4 Issue IX, September 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

tagging of event relationships. Additionally, this paper also evaluated sentiment analysis and effects of temporal characteristics of events on the emotion expressed in the text. For instance, whether news about certain events and the sequence of occurrence in time, affects the sentiment and polarity of emotions expressed in the text. Several multimodal emotion recognition studies have focussed on audio-visual emotion recognition. This paper examines the emotions from text based news feeds and the effect of chronology of events invoking emotions such as rebellions, accidents, coups, natural disasters, calamities, political events and social events.

II. METHOD

For the purpose of temporal analysis of news feeds, data from various popular news websites was gathered. Specifically, political events articles were chosen. A total of 862 articles containing 600 to 1000 words were recorded. The articles were annotated using event linking and manual graph development. This data served as the ground truth. The automatic linking process was dependent on correct tokenization and action verb identification. For tokenization special character and space based delimiters were used. After the tokenization each word token was tagged with properties such as noun, verb, adverb and adjective. Additionally, the word frequency and the location of words from each action verb was measured. The list of temporal words used for position based analysis is shown in Table. 1. The weighted graph was constructed with each event as the node and the number of hops taken to reach the node as the weight of the link. To determine the time sequence between the events the phrases from the temporal word list were used and the pointer to the event was moved in the paragraph based on the occurrence of the phrase in the sentence and a hop count in both directions was maintained with the current timestamp as the reference point. To establish the sequencing of events on a time series the events were ordered by the number of hops taken to reach the words from the current timestamp.

Table. 1 Temporal words

Temporal word list used for phrase detection		
Before	between	shortly after that
After	by	then
later	during	henceforth
yesterday	earlier	tomorrow
Back then	eventually	lately
ago	except	next
same time	finally	next week
simultaneously	following	suddenly
at that moment	for	in addition
at first	from then on	not a moment too soon
at last	in the meantime	now
as soon as	in the end	

The event identification was based on action verbs which were used to create nodes for 4 events with the highest frequency in the news feed. The four different categories to group similar events together were natural disaster, accident, politics and sports. First the events were identified using action and noun words such as {won, championship, defeated, series} for sports, {tremor, rocked, shook, earthquake, scale} for disaster, {hit, run, crash, interstate, pile, traffic, light, cops, bumper, hurt} for accident and {election, speech, controversy, run, voter, senate, politician, businessman} for politics were used. After the event specific nodes were identified the linking was done by moving around the paragraph using search and count method to navigate from event to event using temporal words from the list and counting as the search progressed. The count served as the weights for the links and the links with the highest count established the relation type such as before, after, simultaneously or unrelated.

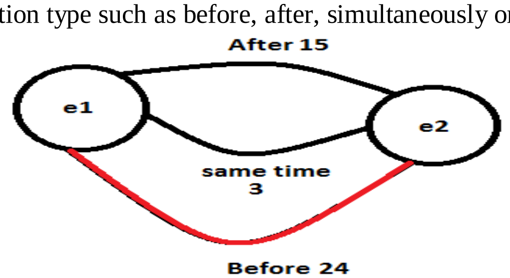


Fig. 1. Weighted event linking based on word hops

Volume 4 Issue IX, September 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

In Fig. 1. The event relationship between e1-e2 was determined to be e1 occurred before e2 because the before category links had the highest weight and hop count. To calculate the weight of the relationship category, the hop count and the time specific word weight was used. Hence the before type relation weight was given by the sum of hops found for before type of links + weight of day, year, month, week and time within the e1 and e2 occurrences. For each different location of e1 and e2 in the entire news feed the link weight was calculated. The average of the weights was used to determine the overall link weight. Once the relationship was determined between various events then the events were allocated in chronological order. In case of conflicts, the relation with higher weight was chosen as the final relation type.

The same process was repeated in both directions and for a range of words with the event as the reference point. The graph based relationship type linking was also done for emotion category recognition between the events. To establish co-relation between the various news feeds and the corresponding emotions, manual annotation was done by 5 annotators. For instance, the news about natural disaster was more likely to invoked sadness and fear. News about a sports championship win was more likely to invoke happiness for the local news feed supporting the winning team. The news on politics was likely to invoke anger, disgust or surprise.

An odd number of annotators was used to avoid contradictory classification and avoiding inter-annotator disagreement. For emotion analysis, the six basic emotions were used: Angry, sad, happy, afraid, disgust, surprise and neutral for fact based news reporting. The next section discusses various metrics such as accuracy comparison, hop count comparison, intra-event emotion category recognition.

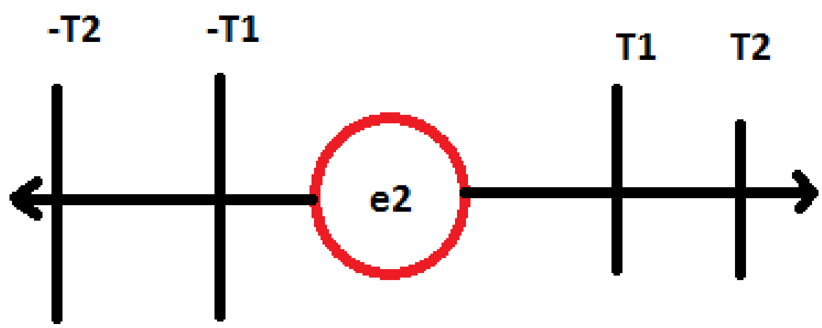


Fig. 2. Bi-directional scan for phrase position

III.RESULT

The results below show the assigned weights for hops because of know time specific words.

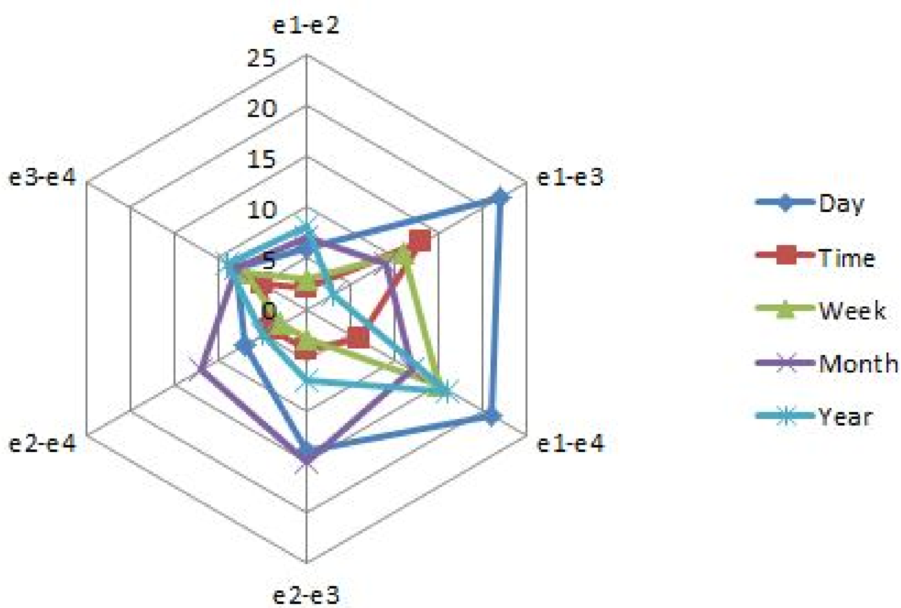


Fig. 3. Weighted hops between event i to j

Volume 4 Issue IX, September 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

The results below show the comparison between weighted hops on the time scale because of phrase positions. The temporal relation scans for the events related by direct dates in the news articles was the highest between e1-e3 and e1-e4. The event relationship between e1-e4 was well defined with day, year and week specific words existing mostly for this relation. Based on the direct time specific words there was no clear consistency between the weights for the event relationships. This indicated that a wider range of temporal word analysis needed to be performed using non-time specific words.

The overall hops found for time-specific words was low compared to the available words from the 800+ news feed articles with each containing 1000s of words. This was another indicator that there was a need to use phrases instead of specific words that describe time for a more accurate temporal event relationship prediction.

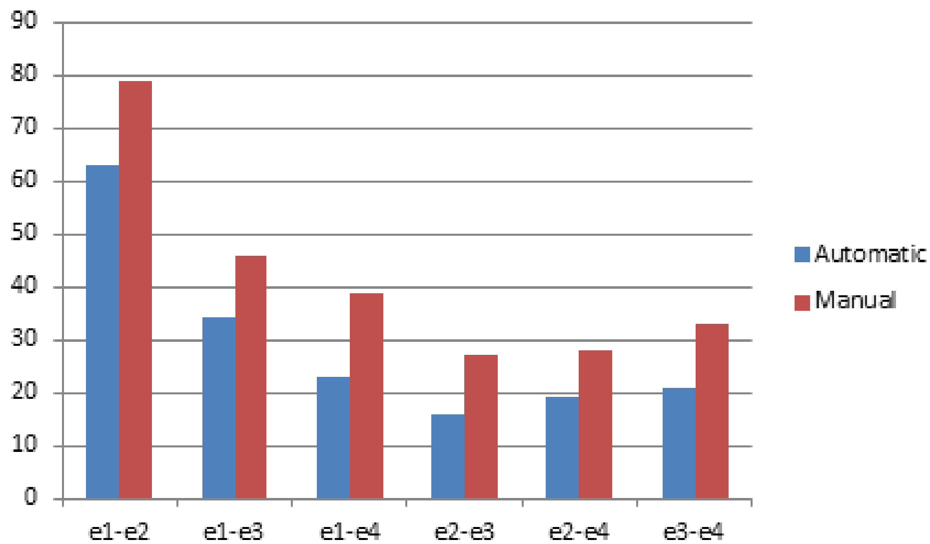


Fig. 4. Automatic vs manual relation identification performance.

The automatic temporal relation detection recall rate was 79.74, 73.91, 58.97, 59.25, 67.85 for e1-e2, e1-e3, e1-e4, e2-e3, e3-e4 respectively for an average of 67% recall rate. Fig, 4 shows the average hop count using various relation type qualifier words for the 6 event pairs between e1 and e4.

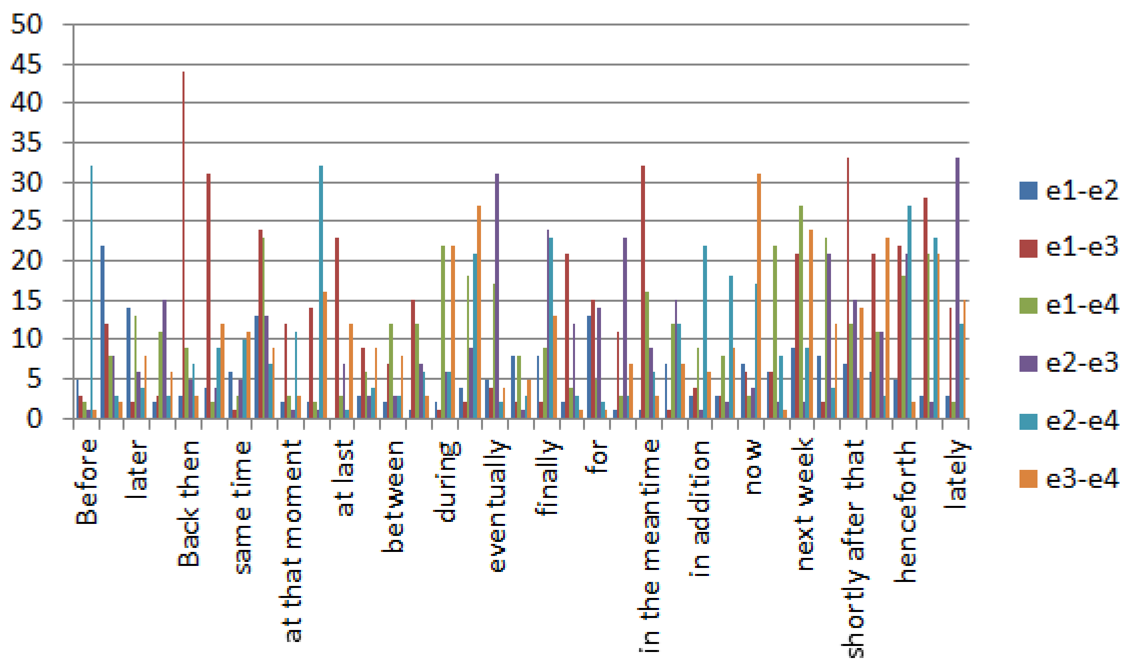


Fig. 5. Hop count by phrase position

Volume 4 Issue IX, September 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Some of the hop count were above 30 which indicated that the news feeds contained two events far apart in the text. One advantage of the phrase postion weighted graph based technique was that even if the hop count was high, it was weighted so the processing was instantaneous with minimum traversal in subsequent tree traversals. There was no clear threshold or pattern among the weights of the temporal phrase based hops. As a result, the various phrases and their hops had to be evaluated separately instead of grouping them into before, after, same time and unrelated categories.

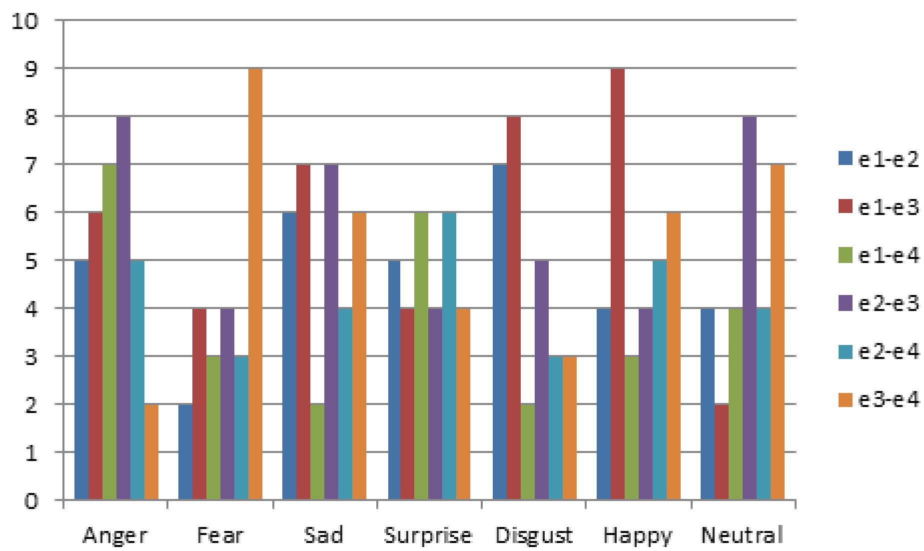


Fig. 6. Emotion occurrence between event relationships

Based on the ranking of emotions found between pairs of events there was no clear threshold to determine the overall emotion expressed in the sentences containing the events across the various event pairs. In case of event e1-e3 happiness was identified as the highest ranking emotion, which indicates that the news was most likely about a sports championship victory. Similarly, fear was identified as the prominent emotion for the event e3-e4 pair indicating the news was most likely related to a natural disaster.

IV. CONCLUSIONS

The automatic estimation of temporal relationship between action verbs from the news feed showed 67% accuracy in terms of recall rate. The weighted-frequency based linked graphical models of phrase positions allowed the automatic estimation to travel bi-directionally while searching for occurrence of events before, after or at the same time as the input event. The estimation performance and processing time was merely seconds, thus making the automatic time sequence prediction steps feasible for real time applications. One such application that we plan to work as future scope is recommender system based on time sequencing where based on the search keyword, the system could suggest news articles related to the event in chronological order. Additionally, the emotion categorization of the event pairs based on the relationship type and the weighted graph traversal technique showed that sentiment analysis could also be performed on the text using the same technique.

REFERENCES

[1] M. Tatu, and M. Srikanth, "Experiments with reasoning for temporal relations between events," In Proc. Coling 2008, Manchester, UK, Aug. 18-22, 2008, pp.857-864.

[2] A. Elkhlifi, and R. Faiz, "Event extraction approach for Web 2.0," aiccsa, pp.1-8, ACS/IEEE International Conference on Computer Systems and Applications - AICCSA 2010.

[3] H. Llorens, E. Saquete, and B. Navarro-Colorado, "TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2", In Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval '10, Stroudsburg, PA, USA, 2010, pp.284-291.

[4] S. A. Mirroshandel, G. Ghassem-Sani, and M. Khayyamian, "Using syntactic-based kernels for classifying temporal relations". Journal of computer science and technology 26(1), Jan. 2011, pp.68-80.

[5] J. Pustejovsky, R. Hanks, A. Sauri, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo, "The TIMEBANK corpus", In Proceedings of Corpus Linguistics 2003, Lancaster, pp.647–656, 2003.

[6] J. F. Allen, "Maintaining knowledge about temporal intervals", Communications of ACM, vol. 26, num, 11, pp. 832–843, 1983.

[7] M. Verhagen, R.Sauri, T.Caselli, and J.Pustejovsky, "Semeval-2010 task 13: Tempeval-2", In Proceedings of the 5th International Workshop on Semantic Evaluation, Stroudsburg, PA, USA, pp.57–62, 2010.

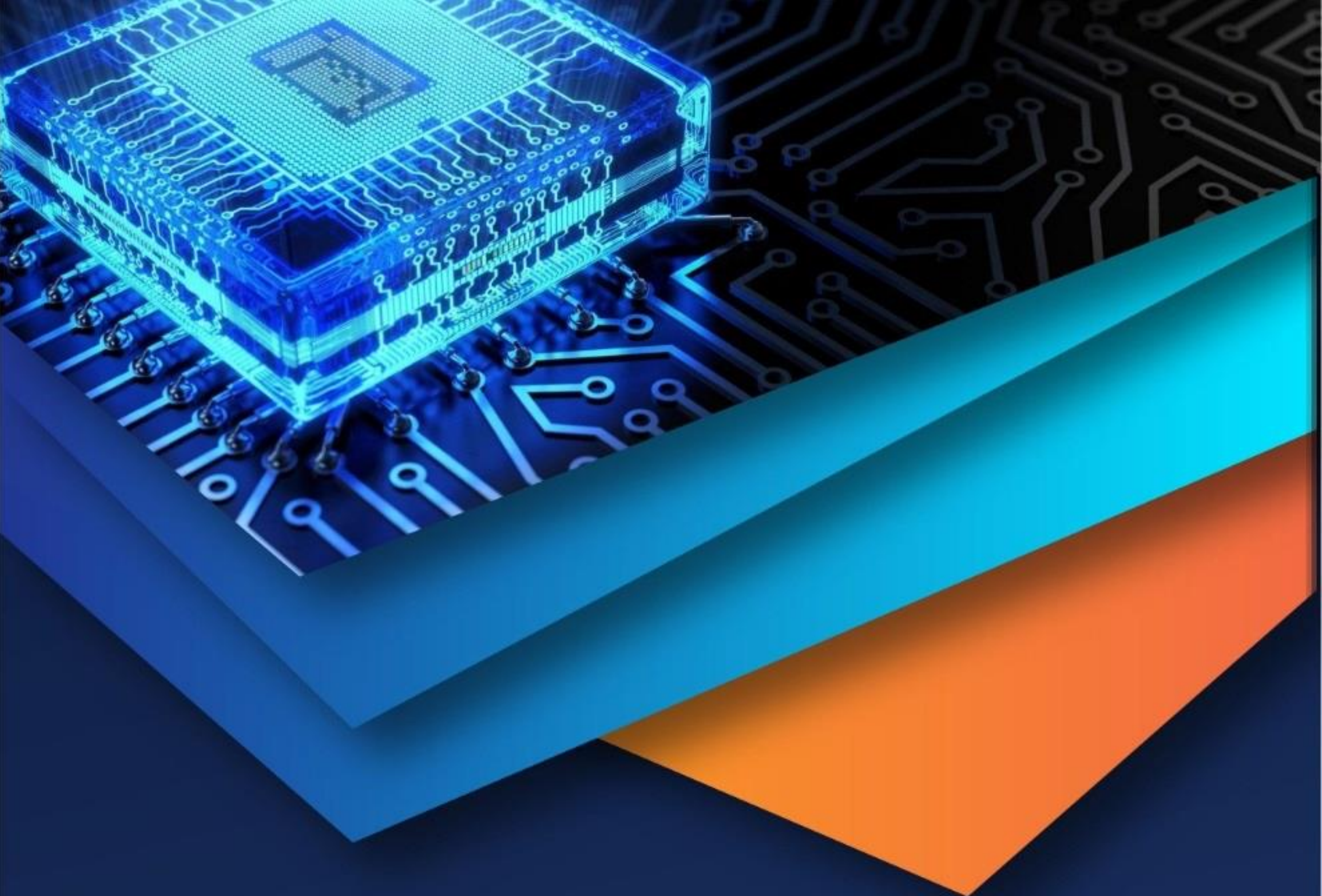
[8] E. Ha, A. Baikadi, C. Licata, and J. Lester, "NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology" In Proceedings of the 5th

Volume 4 Issue IX, September 2016
ISSN: 2321-9653

International Workshop on Semantic Evaluation, SemEval-2010, Stroudsburg, PA, USA , pp.341– 344, 2010.
[9] T. Chklovski, and P. Pantel, "Verbocean: Mining the web for fine-grained semantic verb relations", In Dekang Lin and Dekai Wu, editors Proceedings of EMNLP 2004, pp 33–40, Barcelona, Spain, 2004.
[10] J. R. Quinlan, "Induction of decision trees", Machine Leaning, vol. 1, Kluwer Academic Publishers, 1986.
[11] J. R. Quinlan, "C45: Programs for Machine Learning", Morgan Kaufmann, San Mateo, CA, 1993.
[12] NIST, "The ACE 2007 Evaluation Plan", National Institute of Standardsand Technology, 2007.
[13] P. Kingsbury, and M. Palmer, "From treebank to propbank", In Third International Conference on Language Resources and Evaluation, LREC-02. 2002.
[14] A, Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Drunken Abnormal Human Gait Detection using Sensors", Computer Science and Emerging Research Journal, vol 1, 2013.
[15] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Fear Detection with Background Subtraction from RGB-D data", Computer Science and Emerging Research Journal, vol 1, 2013.
[16] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Code Definition Analysis for Call Graph Generation", Computer Science and Emerging Research Journal, vol 1, 2013.
[17] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Multi-View Point Drowsiness and Fatigue Detection", Computer Science and Emerging Research Journal, vol 2, 2014.
[18] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Group Emotion Detection using Edge Detecttion Mesh Analysis", Computer Science and Emerging Research Journal, vol 2, 2014.
[19] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Polarity Analysis of Restaurant Review Comment Board", Computer Science and Emerging Research Journal, vol 2, 2014.
[20] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Sentiment Analysis in Code Review Comments", Computer Science and Emerging Research Journal, vol 3, 2015.
[21] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Temporal Analysis of News Feed Using Phrase Position", Computer Science and Emerging Research Journal, vol 3, 2015.
[22] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Decision Rule Driven Human Activity Recognition", Computer Science and Emerging Research Journal, vol 3, 2015.
[23] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Depression and Sadness Recognition in Closed Spaces", Computer Science and Emerging Research Journal, vol 4, 2016.
[24] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Dynamic Probabilistic Network Based Human Action Recognition", Computer Science and Emerging Research Journal, vol 4, 2016.
[25] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Fight and Aggression Recognition using Depth and Motion Data", Computer Science and Emerging Research Journal, vol 4, 2016.
[26] A. Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, "Sensor Tracked Points and HMM Based Classifier for Human Action Recognition", Computer Science and Emerging Research Journal, vol 5, 2016.
[27] A. S. Patwardhan, 2016. "Structured Unit Testable Templated Code for Efficient Code Review Process", PeerJ Computer Science (in review), 2016.
[28] A. S. Patwardhan, and R. S. Patwardhan, "XML Entity Architecture for Efficient Software Integration", International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 4, no. 6, June 2016.
[29] A. S. Patwardhan and G. M. Knapp, "Affect Intensity Estimation Using Multiple Modalities," Florida Artificial Intelligence Research Society Conference, May. 2014.
[30] A. S. Patwardhan, R. S. Patwardhan, and S. S. Vartak, "Self-Contained Cross-Cutting Pipeline Software Architecture," International Research Journal of Engineering and Technology (IRJET), vol. 3, no. 5, May. 2016.
[31] A. S. Patwardhan, "An Architecture for Adaptive Real Time Communication with Embedded Devices," LSU, 2006.
[32] A. S. Patwardhan and G. M. Knapp, "Multimodal Affect Analysis for Product Feedback Assessment," IIE Annual Conference. Proceedings. Institute of Industrial Engineers-Publisher, 2013.
[33] A. S. Patwardhan and G. M. Knapp, "Aggressive Action and Anger Detection from Multiple Modalities using Kinect", submitted to ACM Transactions on Intelligent Systems and Technology (ACM TIST) (in review).
[34] A. S. Patwardhan and G. M. Knapp, "EmoFit: Affect Monitoring System for Sedentary Jobs," preprint, arXiv.org, 2016.
[35] A. S. Patwardhan, J. Kidd, T. Urena and A. Rajagopalan, "Embracing Agile methodology during DevOps Developer Internship Program", IEEE Software (in review), 2016.
[36] A. S. Patwardhan, "Analysis of Software Delivery Process Shortcomings and Architectural Pitfalls", PeerJ Computer Science (in review), 2016.
[37] A. S. Patwardhan, "Multimodal Affect Recognition using Kinect", ACM TIST (in review), 2016.
[38] A. S. Patwardhan, "Augmenting Supervised Emotion Recognition with Rule-Based Decision Model", IEEE TAC (in review), 2016.
[39] A. S. Patwardhan, Jacob Badeaux, Siavash, G. M. Knapp, "Automated Prediction of Temporal Relations", Technical Report. 2014.
[40] A. S. Patwardhan, "Edge Based Grid Super-Imposition for Crowd Emotion Recognition", International Research Journal of Engineering and Technology (IRJET), May. 2010.
[41] A. S. Patwardhan, "Human Activity Recognition Using Temporal Frame Decision Rule Extraction", International Research Journal of Engineering and Technology (IRJET), May. 2010.
[42] A. S. Patwardhan, "Low Morale, Depressed and Sad State Recognition in Confined Spaces", International Research Journal of Engineering and Technology (IRJET), May. 2011.
[43] A. S. Patwardhan, "View Independent Drowsy Behavior and Tiredness Detection", International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2011.
[44] A. S. Patwardhan, "Sensor Based Human Gait Recognition for Drunk State", International Journal for Research in Applied Science and Engineering

Volume 4 Issue IX, September 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Technology (IJRASET), May. 2012.
[45] A. S. Patwardhan, “Background Removal Using RGB-D data for Fright Recognition”, International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2012.
[46] A. S. Patwardhan, “Depth and Movement Data Analysis for Fight Detection”, International Research Journal of Engineering and Technology (IRJET), May. 2013.
[47] A. S. Patwardhan, “Human Action Recognition Classification using HMM and Movement Tracking”, International Research Journal of Engineering and Technology (IRJET), May. 2013.
[48] A. S. Patwardhan, “Feedback and Emotion Polarity Extraction from Online Reviewer sites”, International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2014.
[49] A. S. Patwardhan, “Call Tree Detection Using Source Code Syntax Analysis”, International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2014.
[50] A. S. Patwardhan, “Walking, Lifting, Standing Activity Recognition using Probabilistic Networks”, International Research Journal of Engineering and Technology (IRJET), May. 2015.
[51] A. S. Patwardhan, “Online News Article Temporal Phrase Extraction for Causal Linking”, International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2015.
[52] A. S. Patwardhan, “Online Comment Processing for Sentiment Extraction”, International Journal for Research in Applied Science and Engineering Technology (IJRASET), May. 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY