ANALYSIS OF SOCIAL MEDIA TEXTUAL CONTENT USING ACCIDENT DATA SETS FOR CONTEXT RECOGNITION BY GENETIC ALGORITHM

Rashmi H Patil
Department of Computer Science
Rani Chennamma University, Belagavi, India

Siddu P Algur
Department of Computer Science
Rani Chennamma University, Belagavi, India

ABSTRACT

There is huge amount of data which is being processed daily. According to statistic world’s population is 7 billion and 6 billion people has smart phones. So, having smart phones there are various application which connects world. It has made world easier and difficult at the same time. Social media has made people to converse with each other and add few knowledgeable information on it. This huge exploration of data has to be maintained and it is necessary to know the context in which it is used. Discourse interpretation of the word is very important a particular world can be used in positive context or in negative context. For example “great” can be used in positive context as well as negative context. So Natural Language processor is used (NLP) to get the context of the word which is being used. So in this paper we can exploit the knowledge of Natural language processing the context in greater way.

Keywords: discourse, social context, adjectives, perception, elocutionary act, Illocutionary act, perlocutionary act.

Cite this Article: Rashmi H Patil and Siddu P Algur, Analysis of Social Media Textual Content Using Accident Data Sets for Context Recognition by Genetic Algorithm, International Journal of Computer Engineering and Technology, 10(2), 2019, pp. 116-124.

http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=10&IType=2

1. INTRODUCTION

In current generation social media has impacted more on people’s life. They can converse with their friends or relatives in fraction of second with inexpensive way. We can share most of the information people can post status; post resources etc on social media they can be structured or unstructured way. When data is being processed it has to be updated in timely
manner on social media it. Data on youtube, facebook, twitter, whatsapp all the data which is present on the social media has to be updated. For example if person post image or status on social media it has to be provided to the intended user at the same time there should not be extra delay attached to it.

Most of the conversation which you converse with your friends are not in structured way we use slangs to address each other that might not be there in dictionary. And we may mix up two or three language and speak this conversation does not follow any set of rules so it may be considered as unstructured data. With conversation as such there can come typographic error and misunderstanding can increase system because it is not synchronous process. For example, Person can send a message at 10 A.M in one mind set and other user can see the message at night 10 P.M things might have changed time so he could misunderstand whole context since it is asynchronous process.

Some words have different meanings when they are used in different places. Example “kite “it is a bird and as well as toy which has light weight frame attached to long string. So NLP has to be careful when they are dealing with such words. Words must be categorized into spoken and scripts. Some words which might be spoken may not be in the dictionary of the language. Some slang which is often used by the people may not be scripted in the syntax of language for social media has two process lexical semantics and syntax analysis. So this paper tells the various process which is used by Natural Language Processor to distinguish the context in which particular word is used.

2. CONTEXT SPECIFICATION

Context is referred as the event that is taking place in particular instant. This is one of the definitions there can be various definition. There are various types of events on which the context depends on some of them are:

**Discourse:** The context of words in which the word is used may depend on the earlier instant. So, we need to be careful while we use such kind of context because things might have changed after few minutes. For example “There are few grapes on creeper” and “I want it”.

Particular context tells “I want it “this asks the question what you want that is answered by the previous sentence “There are few grapes on the creeper”. So interdependency of two context is known here.

**Common Domain:** Both speaker and hearer should have common interest and common domain. Whatever speaker says hearer has to relate to it else hearer will think in his own perceptive and there might occur conflicts of thoughts.

**Spatio-temporal quality:** This are the sentence which depends on both space and time at which the word is used. For example, news channel they give news at particular instance.

**Physical and perceptual context:** The user’s environment may contain many objects the words which is uttered by seeing those objects has to be considered. Even user’s perception must be considered at the same time

**Social Context:** This specifies the way in which people are associated with each other. If people are conversing with their friends they can have informal talks such as usage of slangs. But when they are speaking to some authority they need to converse formally.

The words can be divided in the sentence and can be examined separately but this causes problem. For example:” I am greatly disappointed”. If we parse the sentence by separating each word conflict might occur. Here “greatly” is positive word but used in negative context. If system parses each word it might think it as a positive context and misinterpret the context of the sentence. So better to interpret the meaning of whole sentence at once.
3. WORDS AND SENTENCES
First when we have sentence, we need to see whether the sentence is following the syntax of
the language. It must follow the syntax else it becomes difficult for the system to interpret the
context in which it is used. Usage of adjectives can either intensify or down tone the situation.
Example: It was very difficult paper. Here “very” is an adjective it is used to tell how difficult
paper was. So adjective can play a major role in describing context of the particular word.
The words can be either scanned in top-down or bottom-up approach.
Example:
It is too sunny outside not good enough to play outside.
In above context the words good enough if you take it basically specifies positive context. So
you need to take in positive context and negate it in order to make it as negative sentence in
order to define above context.
This restaurant proposes good quality Greek cuisine in a warm atmosphere
Here everything is in positive context. It tells it provides good quality of Greek cuisine in
warm atmosphere it doesn’t have any negative words in it. This can be used as positive
message.
Adjectives in different context
There is a long queue.
The smart phone has a long battery life.
Here the adjective “long” is used in two different contexts. First long refers to large number of
people waiting for the queue. In example4, it tells smart phones can work without dying for
long duration of time. In example3, long refers to large number of people where as in
example4 it refers to duration of time. Some times in social media we tend to specify the word
by repeating letters of the word to show excitement of current situation that is happening or
happened previously.
This mobile has amazing feature
It is the sentence.
People often write in this context.
This mobile has amaazzzzzinnngggg featuresss.

4. DISCOURSE STRUCTURE
Words and sentence have proper format in the language. They cant be arranged as our wish. If
words are misplaced then system cant understand the syntax. Words in sentence are cohesive
in nature. Words cant be separately read and found out the meaning for it. If we do system can
completely misinterpret the context in which it is used .So we can conclude that words in
sentence has cohesive nature.
There are basically two approaches to analyse discourse of context
- Dynamic semantic
- Theory of discourse structure.

4.1. Dynamic semantic
When S is a sentence which is used in current situation and it may or may not depend on the
earlier context. If it depends on the previous context it is necessary for us to know the earlier
context else system might be in dilemma.
Example:
K: There is a file
S: I want it
S is the sentence in current situation “it” is unknown. System can misinterpret if it doesn’t know the previous context so it is necessary for the system to know the previous situation as well

4.2. Theory of discourse
It links discourse meaning from explanation, elaboration, narration. Discourse relation can be derived from those links. By using such links we can make out the writer perceptive what were their intension to write such columns etc. A number of theory has been raised for to know the meaning of different context

Example:
[Geetha had been to restaurant](stat1) [ambience was really good](stat2) [she ordered starter and main course](stat3) [main course was good but starter was not that good]

From the above example you can find there are many sentences work concurrently to explain the current situation. stat1 refers to the speaker has been to one restaurant. stat2 refers that the surroundings was really nice. Till this you find everything is normal and is in positive context. So system may assume that this is positive context and speaker is happy about the restaurant but in stat3 pronoun is used instead of noun so inorder to understand stat3 stat1 should be known there speaker specify who goes to restaurant. In stat4 speaker specifies that she is unhappy with the starters so model has to see whole context to understand the perspective of user. So each and every sentence in the clause makes sense and has to be specified.

5. BEYOND WHAT IS SAID?
We should mainly classify between what is uttered and what social implication did that cause after uttering those words. So basically there are three perception in which context can be described.

- Elocutionary act: It defines the perception of speaker in which context they uttered those words and what they might expect from the hearer.
- Illocutionary act: It defines out of what circumstances or situation did those were uttered from the speaker. It might differ from one situation to another. Speaker can order or request so it is models duty to understand the context well.
- Elocutionary act: It defines the hearer’s perceptive in which they take speaker’s word. For example hearer may not like speaker so even if they request it might seem like commanding. So pragmatic analysis has to be done carefully.

Speech acts are the semantic/pragmatic counterpart of sentence types. The sentences types affirmative, interrogative, and exclamative correlate with the speech acts of assertion, question, expression.

For example:
- I want that file at any cost.

Model may interpret as requesting but it is commanding. So whole interpretation must be done again. Model can find the intension of users by asking question to them. Answers which is given by makes the model to predict the user’s intention towards the particular instant.
Example (1):
- Do you want pizza or pasta?
- Answer: I prefer none.
  This tells that user prefers none and she/he might not be interested in those kind of food.

Example (2):
User 1: Shall we go to movie
User2: I have lot of work to do

This example shows what user 1 wanted and user2 has denied that. User 1 was requesting user to go out for movie but user two was not interested in movie so denied the proposal.

Similarly, in the context of accident term as the context, we could able to identify the context in the same manner, which is shown below:

- Women are more likely to get in a car accident on a weekday, men more likely to crash on the weekend.
- Police Take Second Look At Venus Williams’ Car Accident With New Footage
- Looks like someone had a shaving accident in St Bees, Cumbria! Sea foam blown onshore by thanks to
- In California, car accident claims are processed digitally through remote sensing and drones.
- The way fat people always play the victim you’d think they ate all that food by accident
- I’ve probably made about 200 calls on snapchat n every single one was a fucking accident.
- That’s six months after Argwings-Kodhek is killed in a staged accident. Then Russia Hospital follows.
- When there is an accident the driver is held responsible. ATA represents businesses not drivers.
- Ed Sheeran cancels Asia tour dates after cycling accident
- Ed Sheeran ‘unable to perform live concerts for the immediate future’, postpones some Asia tour shows
- Fake News: Mr. Bean AKA Rowan Atkinson NOT Dead at 58 After Car Accident
- John Mayer wishes Superman Ed Sheeran well after bike accident
- As many of you know one of our members of the SIU Greek community was injured in an accident this weekend. The...
- Real question for London transportation dept: Will there be a campaign against jihad hate crimes too?
- I updated with a full story of mt accident and medical bills, every donation and retweet is appreciated

6. DISCOURSE PHENOMENON

6.1. Discourse Structure and Coherence Modeling

Although the determining of discourse for traditionally written text is now well established there is little work on applying discourse theories to social media text.

They propose a plan to physically comment on tweets as indicated by Rhetorical Structure Theory standards and found that up to 40% of German tweets are a piece of discussions, and that answer-relations create discourse trees. The analysis of Twitter-specific phenomena
reveals that URLs carry communicative content, talk relations, are once in a while express (just 20% of the cases). They additionally see that causal connectives are visit in Twitter: 1.7% of the tweets and 2.6% of the answers.

6.2. Argumentation mining
In social media lots of opinion will be raised by the user for certain facts. If der is a certain issue there will be lot of opinion raised because all human have there own perspective and has freedom of thinking. All human’s thinking can’t be same. Say if user A raises the same opinion about the fact that is put up on the social media User B can oppose it because his point of view might be different this may cause many arguments on social media.

Arguments and disagreements may just explode the data without getting any conclusion on it. User may deviate from the actual fact and may think in his perception and a comment on it and may live in illusionary world. So binary classifier is used for arguments and disagreements to find the intensions of the user.

6.3. Intention Detection
Intention detection has become hot stream in the market. The intension of the user has to be provided for many system. For example recommender system if it gets to know on what category does the user is interested in, it can provide items of those category. This topic is trending user’s intension can be analysed with mouse click many research can be done on it.

The most of analyse which is taking place is on assertion, recommendation, request, question, comment. Assertion it detects the intension of user by analysing why that question was asked what made user to ask such question. Recommendation it analyse the user intention and gives the recommendation of relative objects. Comments can decide whether the user is negative or positive person.

6.4. Demographic Information
In demographic information it is basically detects age, gender, race, income, location and many others. The users age can be determined by the way they speak and way they behave in social media. As time grows even maturity increases so most of the people who is aged don’t involve in all the arguments going on social media they try to ignore. They speak in formal way they don’t use slangs as teenagers or youth use. Even gender can be easily determined because female does not speak certain way as men do. And there choices, speaking style is different when compared to men. So we can easy distinguish between men and women.

6.5. Social Network
Social network means the circle of friends, relatives, known person which you have in your social media. Most of your friends may have same perspective as you have because to be friends you need to have common goals not same perspective at least similar perspective should be pursued. Similar community may have similar subjective on the fact which are put up on the social media. Most of the trolls happen on social media between two groups of people each group may have common people with common perspective and they might troll other group for no good reason.

7. OTHER ARTICLES OVERVIEW
This issue intended to think about how the treatment of semantic wonders, specifically at the talk level, can benefit NLP-based web-based life frameworks, and help such frameworks advance past portrayals that incorporate just packs of words or sacks of sentences. Talk and
down to business data can likewise help move past sentence level methodologies that ordinarily represent nearby relevant wonders depending on devoted vocabularies and shallow or profound syntactic parsing. All the more imperatively, the point of this issue is to demonstrate that consolidating semantic insights, discourse information, and other logical marvels, in mix with the measurable abuse of information, can result in an enhancement over methodologies that exploit just a single of those viewpoints. We got a sum of 15 entries, reflecting a significant enthusiasm for these marvels in the computational semantics network. After a thorough survey process, we chose six articles, covering different parts of the theme. The chose articles address profound issues in phonetics, computational etymology, and sociology. The uncommon issue is organized around three principle topics, as per the sort of setting considered in each article:

- Social setting: The attention here is on the social and social importance in online discussions from a hypothetical perspective
- Conversation turns and sound judgment learning: Here, we aggregate papers that review wonders for which individuals make inductions in their ordinary use of dialect, concentrating on inductions that are drawn while hunting down the figurative importance of an expression
- Conversational setting: The third part centers around the job of talk wonders in handling online life discussions, including topicality (Li et al.), discourse acts (Joty and Mohiuddin), and argumentation (Cocarascu and Toni).

Whatever remains of this area gives a short prologue to every one of the six acknowledged papers. The article by Kieslingetal researches string structure and semantic properties of taking from the online stage Reddit. Stancetaking catches the speaker's (or writer's) relationship to the subject of exchange, the questioner, or group of onlookers, and the discussion (or composing) itself. The writers first propose another informational collection where discussion strings are clarified by three connected position measurements: influence, venture, and arrangement. These measurements are then anticipated depending on lexical highlights. The quantitative and subjective consequences of this investigation demonstrate that position articulations will in general example in cognizant conversational strings.

"A Joint Model of Conversational Discourses and Latent Topics on Microblogs" separate subjects from microblog messages, a testing assignment given the information sparsity in short messages that frequently need structure and setting. To address this issue, the creators speak to microblog messages as discussion trees dependent on their reposting and reply ingrelation, and propose an unsupervised model that jointly learns word appropriations to distinguish the diverse elements of conversational talk and different idle subjects to speak to content-specific data installed in microblog messages. Their investigations demonstrate that the proposed joint model on subject soundness beat cutting edge models. The yield from the joint model is then utilized for microblog rundown: By moreover catching word circulations for various supposition polarities, the together displayed talk and subject portrayals can successfully demonstrate synopsis commendable substance in microblog discussions. The article by Ghosh et al. ("Mockery Analysis Using Conversation Context") thinks about the job of discussion to distinguish mockery in tweets and exchange forums. The setting considered here concerns the present turn just as the earlier and the succeeding one. Inorder to show to what degree displaying of discussion setting helps in mockery location, the creators research both established learning models with phonetically roused discrete highlights and a few kinds of LSTM systems (restrictive LSTM arrange, LSTM systems with sentence-level consideration). The models were tried on various corpus class informational indexes and the outcomes demonstrate that consideration models accomplish significant enhancement when utilizing the earlier turn as setting for every one of the informational collections. To all the more likely measure the difficulty of the errand, the creators play out a subjective
investigation of consideration loads delivered by the LSTM models and examine the outcomes contrasted and human execution on the undertaking. In the article by Van Hee et al. ("We Usually Don't Like Going to the Dentist: Using Common Sense to Detect Irony on Twitter"), the job of setting in figurative dialect location is additionally investigated. Contrasted and Ghosh et al., who centre around conversational setting, Van Hee et al. target presence of mind and demonstrative learning and propose to show verifiable or prototypical opinion (e.g., "flight delays," "heading off to the dental specialist" for the most part pass on negative assessment) in the system of programmed incongruity discovery in tweets. Their methodology utilizes a help vector machine classifier depending on lexical syntactic, and semantic features, with a particular focus on lexical and semantic features that have been extended with language model features and word cluster information. The results show that applying sentiment analysis using Sentic Net and real-time crawled tweets is a viable method to determine the implicit sentiment related to that concept or situation. Cocarascu and Toni (Combining Deep Learning and Argumentative Reasoning for the Analysis of Social Media Textual Content Using Small Data Sets) propose a method to check whether news headlines support statements from tweets, to allow for fact-checking. Their deep learning method extracts argumentative relations of attack and support. Then they use the proposed method to extract bipolar argumentation frameworks from reviews, to help detect whether they are deceptive. They show experimentally that the method perform swell in both settings. In particular, in the case of deception detection, the method contributes a novel argumentative feature that, when used in combination with other features in standard supervised classifiers, outperforms the latter even on small data sets. The last article in this special issue, by Joty and Mohiuddin (Modeling Speech Acts in Asynchronous Conversations: A Neural-CRF Approach), presents a method for speech act recognition, a problem that has long been a concern in the spoken dialogue research community, and one that poses particular problems in online social media communication, which tends to be offbeat. Joty and Mohiuddin train LSTM RNN sussing conversational word embeddings. This is a significant result, as they demonstrate that word embeddings prepared on a related area enhance the execution of the framework. The commitment of this article is to fuse setting as conditions crosswise over sentences. It is obvious from the writing that discussion structure is significant when deciphering discourse acts. The creators propose to demonstrate it as a chart structure, given the nonlinear idea of nonconcurrent discussion. Furthermore, Joty and Mohiuddin work from the theory that, while speaking to sentence significance, word arrange is critical, and ought to be protected. In spite of the fact that this does not appear to be a progressive idea, word arrange is regularly dismissed in "great" machine learning approaches, and in present day vector portrayals of content.

8. CONCLUSIONS
From all the study made above I conclude that there are lot more to be done on this field. Because data is being exploded day by day because of smart phones and most of the smart phones has social media on it most of the people are on social media so its necessary to know the context of message which is put on the social media. And what is the intention of people who is posting the message it is necessary for Natural language processor to create new model to handle all these stream without misinterpreting the data. To detect all this we need to develop a tool which detects perspective of user in right way. There are many tools that has been developed but some are victim of pragmatic error. If tool is developed they can be used to find sentiment of the user and user choice or many more. All this can be used for many systems which relay on data examples recommender system they take user interest and give related objects so it will be user to view the objects as well as company to gain profit.
9. FUTURE WORK
The research can go in various directions such as interactions between content and extra-linguistic or extra-textual features, in particular time, place, author profiles, demographic information, conversation thread, and network structure.

REFERENCES

[1] Farah Benamara, Diana Inkpen, Maite Taboada “Introduction to the Special Issue on Language in Social Media: Exploiting Discourse and Other Contextual Information”, 10 September 2018. doi:10.1162/coli a 00333, Association for Computational Linguistics. Addawood, Aseel and Masooda Bashir. 2016.

[2] “What is your evidence?” A study of controversial topics on social media. In Proceedings of the Third Workshop on Argument Mining, ArgMining 2016, pages 1–11, Berlin, Germany.

[3] Aiello, Luca Maria, Georgios Petkos, Carlos J. Martín, David Corney, Symeon Papadopoulos, Ryan Skraba, Ayse G’oker, Ioannis Kompatsiaris, and Alejandro Jaimes. 2013. Sensing trending topics in Twitter. IEEE Transaction of Multimedia, 15(6):1268–1282.

[4] Allen, J. F. and C. R. Perrault. 1980. Analyzing intention in utterances. Artificial Intelligence, 15(3):143–178.

[5] Allen, Kelsey, Giuseppe Carenini, and Raymond T. Ng. 2014. Detecting disagreement in conversations using pseudo-monologic rhetorical structure. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, pages 1169–1180, Doha.

[6] Asher, Nicholas and Alex Lascarides. 2003. Logics of Conversation. Cambridge University Press. Attardo, Salvatore. 2000. Irony as relevant inappropriateness. Journal of Pragmatics, 32(6):793–826.

[7] Austin, John Langshaw. 1962. How to Do Things with Words. Oxford. Bach, Kent. 1997. The semantics-pragmatics distinction: What it is and why it matters. VS Verlag für Sozialwissenschaften. pages 33–50.

[8] Baly, Ramy, Mitra Mohtarami, James R. Glass, Lluís M’arquez, Alessandro Moschitti, and Preslav Nakov. 2018. Integrating stance detection and fact checking in a unified corpus. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 21–27, New Orleans, LA.

[9] Bamman, David and Noah A. Smith. 2015. Contextualized sarcasm detection on Twitter. In Proceedings of the International Conference on Web and Social Media, ICWSM 2015, pages 574–577, Oxford, UK.

[10] Barzilay, Regina and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. Computational Linguistics, 34(1):1–34.

[11] Benamara, Farah, Nicholas Asher, Yannick Mathieu, Vladimir Popescu, and Baptiste Chardon. 2016. Evaluation in Discourse: a Corpus-Based Study. Dialogue and Discourse, 7(1):1–49.