Research Abstract(Mine for MITACS)

Presently, I am working on detection of sarcasm in Twitter data. Sarcasm is the special sentiment which infer the opposite meaning of what people trying to convey in the text. It often expressed using positive or intensified positive words. Nowadays, posting sarcastic messages in social media like Twitter, Facebook, WhatsApp, etc. became a new trend. In the presence of sarcasm, sentiment analysis on these social media texts are becoming more challenging task. In this work, I am associated with Mr. Santosh Kumar Bharti ((<https://www.researchgate.net/profile/Santosh_Bharti3>), Ph. D. scholar in National institute of Technology, Rourkela, India.

# Our current work includes validation of tweets using finite state machine (FSM) and then identification of sarcasm in tweets based on certain #hashtags and its context. We have developed an FSM model for tweet validation then classified sarcasm in tweets based on their occurrences. Detailed can find in research paper entitled “Sarcastic Sentiment Detection in Tweets Streamed in Real time: A Big Data Approach”. Further, we will extend the FSM model to detect sarcasm in tweets. In future, we will target some low resource languages for sarcasm detection such as Indian language – Hindi, Telugu, Tamil, etc.

Abstract(for paper)

In this world there are plethora of data. Data from your school records, data of your office attendance, data of your computer memory utilisation etc. In recent years’ methods had been found and new methods are being discovered to use this data in constructive manner where constructive means to use for the benefits of society. The idea of all these processes can be obtained from one of the popular and mind boggling trend “collecting openly available data and analysing data” followed by research community. Twitter has evolved as the first choice for researcher to collect data and use them in their activities. Not only research community but commercial community also using these data to build products. These products are changing lives of millions. Due this practice followed by societies it becomes tremendously important to have meaningful message on the platforms from which data is going to be collected.

By allowing users from all over the world these communities open space for bigger sample space which is an advantage. But with this comes a disadvantage of having messages with no meaning or having meaning not related to the field. Having those messages on the sample space increases time and space required for the intended result. Our work is to define messages on these platforms so that we can discard such messages from sample space and save time.