**Mining Social Media with Social Theories: A Survey**

The purpose of this paper is to show how social media mining is different from traditional data mining, and that social media data is more social, has lots of gaps, and is highly linked and unstructured. The author proposes that these problems can be addressed by combining social science theories and the algorithms of computer science. The problem on computing social media data is that it has unstructured data such as short text, slangs, typos, spacing errors etc. Another problem with social media data is that it is incomplete and makes it difficult to perform analysis on the data. Social theories propose by the author in the paper are: (1)Social correlation theory like: Homophily, Confounding, Influence. (2)Balance theory: states that how two users having a common node are connected together in the social network. (3)Status Theory: states that rank of a user in the social network. Social theories can be used together to form a huge social matrix which can be used for user-related tasks such as: (A)Community Detection: where users who share similar tags, share and view similar post, have a common set of followers and followees are grouped together in the same community. (B)User Classification: classify a set of users so that appropriate services can be provided to them. (C)Social Spammer Detection: this detects spam accounts by using the fact that normal users behave similarly with their neighbors while spammers don’t. (D)Link prediction: these are used for giving recommendation to users, users in unsigned networks are given recommendation based on homophilly while users in signed networks are given recommendation based on their degree of overlapping content with their neighbors. (E)Sentiment analysis which states similar users exhibit similar sentiments. Future research needs to be done to establish more concrete social theories which can be applied to social media data.

**Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination**

The purpose of this paper is to perform and compare Lexicon methods and Machine Learning methods to perform twitter Sentiment Analysis. However the best results were obtained when both the techniques were combined together, Lexicon score was calculated using Lexicon methods and then using this a model is trained based on machine learning techniques. Earlier researchers used bagging, boosting, random sampling, Bayesian Model Averaging ensemble methods to compute the sentiments in a tweet. The twitter data needs a lot of pre-processing before carrying computing the lexicon score or training a model:(1)Part-of-speech Tagging. (2)Stemming and Lemmatization. (3)Stop-words removal. (4)Negation Handling. (5)But-clauses. (6)Tokenization into N-grams. After preprocessing the polarity of lexicon is checked by using a bag-of-words and then the sentiment score is calculated as the average of all the words polarity. These lexicons can either be manually created or constructed from a trained data. However in Machine learning based approach the after data pre-processing features are generated and selected. These features are then fed into a Learning Algorithm which consist of algorithms like Decision Trees, SVM, or Naïve Bayes. The F-score which is similar to a lexicon score can then be computed on a test dataset. Another strong point about the authors approach is that they considered slangs and emoticons, and manually gave them a polarity and a score. The authors also ignored the grammatical construction in which the negation word did not have any scope. However the weak point was that they used features which had an information gain above 0 for feature selection this feature selection might not perform so well on an unseen dataset. Also the size of the lexicon dictionary was too big which decreased their performance due to ambiguity of words polarity and increased the model complexity.