Evaluation Social Ties and Trust in Online Social Network

[1]Aditi Sharma, [2]AP Nidhi

Department of Computer Science, Swami Vivekanand Institute of Engineering & Technology,
Punjab Technical University, Jalandhar, India

Abstract: ***In this research work we have built a systems which pulls tweets , pre-process each tweet to remove unwanted artifacts and the gives stemming treatment to each token in the tweet to finally get the score value which is calculated on the basis of numerical strength expressed in mathematical expression for intimacy, trust , social distance between the parties interacting on digital media twitter. The nature of relationship with respect their tie is calculated on the bases of multivariate linear regression analysis and to further improve on this method which have conducted polynomial regression for finding a better and more accurate way to data fit the twitter tie dataset which is apparent from the statistical test value of Coefficient of determinant and other statistical test .The values of y-intercepts are interpreted and illustrated in results section .***

Keywords: Social Network Analysis, Regression and Coefficient of determinant

1. Introduction

Now days, the online social networks (OSN) have gained considerable popularity. More and more people use OSN to share their interests and make friends, also the OSN helps users overcome the geographical barriers. Social media treats all users the same; trusted friend or a total stranger with little or nothing in between [1]. However, the relationships lie on a continuum between total stranger and close friend. The present study aims to identify the factors in social media (twitter) that accurately predict trust between the two users.

Relationships make social media social. However, online social media treats all the users the same: trusted friend or total stranger, with little or nothing in between. In reality relationships fall everywhere along this spectrum - total strangers to close friends. Trusted friends and family members can affect emotional health and often join together to lead organizations through tough times. Since the relationships are now becoming predominately online. It is therefore of interest to investigate how well online social media data is able to predict trust among individuals. This is more as important as individuals are not in physical proximity to each other and yet their lives are influenced by the communication they have online end of introduction we propose a predictive model that maps social media data to trust based on following points.

1. Corpus collection of emoticons/text indicating trust (dependent variable), intimacy, intensity, social distance and friendship (independent variable) using twitter API
2. Tweet content analysis of the obtained corpus
3. Obtaining term frequency score for each of the variable (dependent as well as independent)
4. Computing 'In between ratios' for dependent and predictive variables.
5. Establishing relationships using linear as well as Polynomial regression (proposed model)

2. Methodology

Micro blogging today has become a very popular communication tool among Internet users. Millions of users share opinions on different aspects of life every day. Therefore microblogging web-sites are rich sources of data for opinion mining and sentiment analysis. As microblogging has appeared relatively recently, there are a few research works that were devoted to this topic. In this study, we focus on using Twitter, the most popular microblogging platform, for the task of sentiment analysis. We show how to automatically collect a corpus for sentiment analysis and opinion mining purposes. We perform linguistic analysis of the collected corpus and explain discovered phenomena. Using the corpus, we build a sentiment classifier that is able to determine sentiments relating to intimacy, intensity and trust. In our research, we worked with English; however, the proposed technique can be used with any other language.

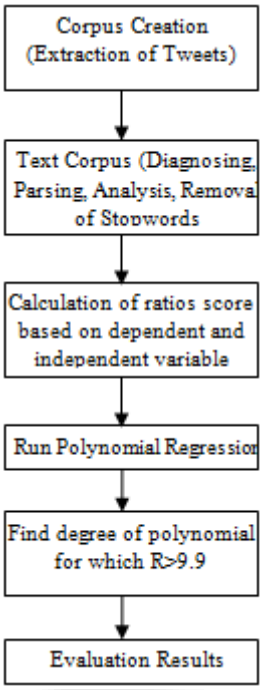
3. Assumptions of the Study

Study was carried out over a period of 15 days (commencing from February 02, 2013 to February 16, 2013) examining at-least six exchanges per day. It was also assumed that if a person used one intimate word for another person, the other person also reciprocated at-least once. Our 35 participants, primarily students came from more than 10 different academic departments. The sample consisted of 26 females (74%) and 9 males (26%) ranging between 18 to 25 years of age.

The methodology followed in this study may be represented as below:

1) Process of corpus collection using twitter API
2) Linguistic analysis of the obtained corpus
3) Obtaining term frequency score
4) Computing 'In between ratios' for dependent and predictive variables.
5) Establishing relationships using Linear and Polynomial regression

4. Flow Graph



5. Implementation

Process of Corpus Collection

Using Twitter API we collected a corpus of text posts and formed a dataset of five classes: representing four emotional states – intimacy, intensity, social interaction and trust; one friendship score [2]. To collect these sentiments, we followed the same procedure as in (Read, 2005; Go et al., 2009). We queried Twitter for three types of emoticons:

- Intimacy related emotions/text.
- Intimacy related emotions/text.
- Trust related emotions/text
- Social Distance

In addition to these emoticons text messages were also classified on the basis of certain words representing these emotions. In order to collect a corpus of objective posts, we retrieved text messages from Twitter accounts of friends who regularly tweet among each other for the purposed of analysis. Because each message cannot exceed 140 characters by the rules of the microblogging platform, it is usually composed of a single sentence. Therefore, we assume that an emoticon within a message represents an emotion for the whole message and all the words of the message are related to this emotion. In our research, we use English language. However, our method can be adapted easily to other languages since Twitter API allows specifying the language of the retrieved posts.

6. Predictive Variables

6.1 Intimacy Variable

Intimacy variable measures the intimacy between two tweeters on the basis of appearance of certain words characterized as intimate words. A dictionary of these intimate words was generated on the basis of Delpi technique. Academicians who were expert in the language of English were asked to list down words depicting intimacy and the words which were common in all the lists were selected.

6.2 Intensity Variables

Intensity variable refers to the frequency of communication between the two friends. It represents the number of words exchanged over the period under study. It also includes the number of times a particular person has appeared in a tweet of another person (tweet length). A dictionary of intensity words was generated in the same way as dictionary of intimacy words using Delphi technique.

6.3 Trust

Trust refers to the faith that tweeters have amongst them. It indicates the extent to which they can confide among themselves and the degree to which they feel comfortable amongst each other. A dictionary of words indicating trust between the tweeters was generated using Delhi technique as explained above.

6.4 Social Distance

The term social distance refers to the educational difference between the two friends. A dictionary of words that indicates the educational difference is prepared in the same way as explained above. The term frequency is calculated and social distance score obtained.

6.5 Friendship Score

Friendship score represents a ratio of incoming friendships to outgoing friendships. Incoming friendship refers to the number of followers a person has and outgoing friendship number of persons a person id following.

7. Results/Graphs

Microblogging now-a-days has become one of the major types of the communication. A recent research has identified it as online word-of-mouth branding (Jansen et al., 2009). The large amount of information contained in microblogging web-sites makes them an attractive source of data for opinion mining and sentiment analysis. The present study attempts to establish relationship of trust that a tweeter may have with other tweeters in his/her contact. In order to understand the relationship we have trust as a dependent variable and intimacy, intensity, social distance and friendship score as independent or predictive variables. A simple and polynomial regression is used to establish the relationship between these.

As a first step in the study we identify the term score for each of these variables (dependent as well as independent). Based upon this term score we compute the scores for these variables. A screen shot of the term score for intimacy is given in Table 1. Similarly term frequency scores for each of the independent/predictive and dependent variable was obtained through twitter API. Based upon these term frequency scores ‘in between ratios’ for all the variables

were obtained. Represented below (Table 2) is the obtained 'in between ratio' for Friend "F1".

Table 2: In Between Ratio for Friend F1

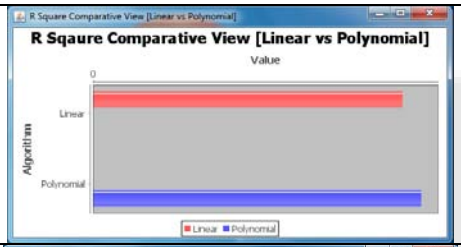
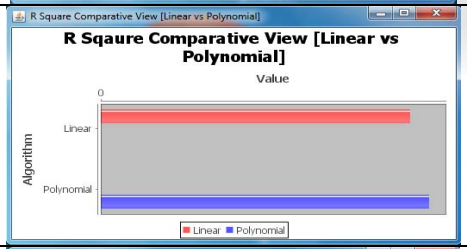
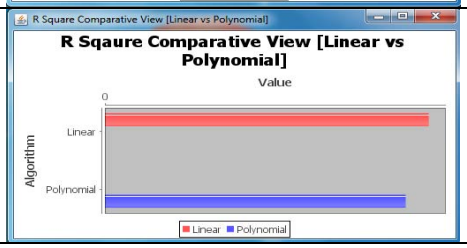
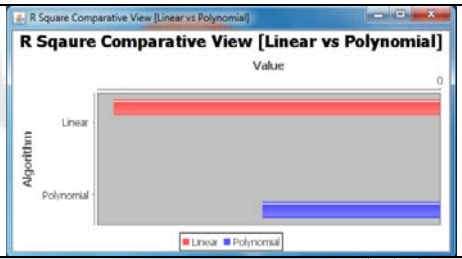
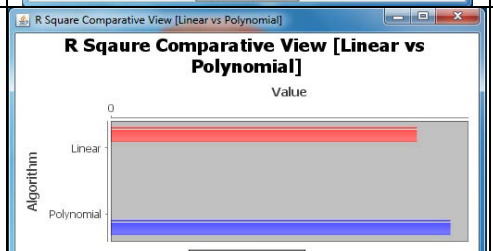
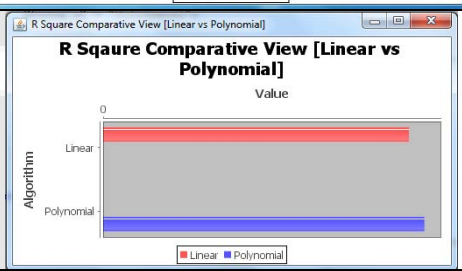
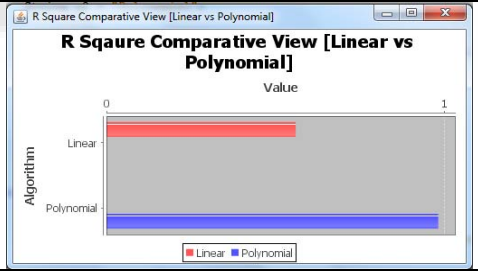
Trust		Intimacy		Intensity		Social Dist.		F. Ship	
F1F2	1	F1F2	2	F1F2	0.4	F1F2	0.6	F1F2	0.9
F1F3	1.3	F1F3	2.2	F1F3	0.4	F1F3	0.9	F1F3	1.1
F1F4	1.6	F1F4	1.5	F1F4	0.6	F1F4	0.6	F1F4	1.5
F1F5	2	F1F5	1.9	F1F5	1.7	F1F5	1.35	F1F5	1.9
F1F6	2.1	F1F6	2	F1F6	1.9	F1F6	1.7	F1F6	2
F1F7	1.8	F1F7	1.75	F1F7	1.9	F1F7	2.1	F1F7	1.75
F1F8	2	F1F8	1.89	F1F8	1.3	F1F8	1.75	F1F8	1.89
F1F9	0.7	F1F9	1.4	F1F9	0.8	F1F9	0.65	F1F9	0.65
F1F10	0.9	F1F10	1.74	F1F10	2.2	F1F10	0.87	F1F10	0.9

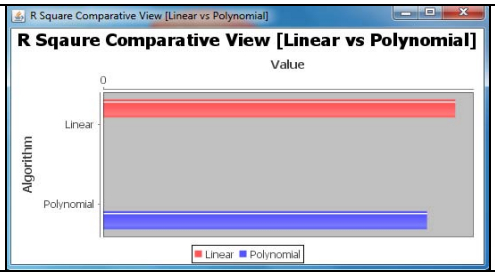
Partial Matrix of variable used for regression

Results obtained after conducting linear regression on the above are presented using graphs which include R^2, Adjusted R^2, co-eff comparative between linear and polynomial regression

8. Graphs

The value 0.56 for R^2 indicates fairly decent relationship between the dependent and independent (predictive variables). Further values of coefficients reveal some interesting information.

1.		Adjusted R Square
2.		Friendship Score
3.		Intensity
4.		Intercept
5.		Intimacy
6.		Multiple R
7.		R Square

8.		Social Score

9. Interpretation of Results

Negative sign indicates a negative relationship and a positive sign indicates a positive relationship. If we assume coefficients of the predictive variables to be zero, meaning there is no intimacy, interaction and friendship (either incoming or outgoing), negative sign of the intercept indicates that friend F1 is distrustful of strangers i.e. with whom he/she has no interaction.

Positive signs for all other coefficients indicate there is a positive relationship between trust and other predictive variables. However, the value of the coefficient represents the contribution of each of these variables to trust. Small positive value for intensity indicates that although it is positively related but doesn't influence trust. In other words high frequency of interaction doesn't necessarily indicate trust in a relationship.

Positive and large coefficients for intimacy and friendship indicate that these two predictive variables have a strong influence on trust. Small positive coefficient for social distance indicates that this variable has relatively little influence on trust. Value of R^2 (0.987875086) as well as Adjusted R^2 (0.978080081) is fairly high indicating that Polynomial regression is better in predicting relationships; it gives a more accurate value of intercept.

10. Conclusion

In this research we examined the use of microblogging for evaluating trust among friends communicating through twitter. The implications of this research include that microblogging is a potentially rich avenue for sentiment analysis. In the present study we have used polynomial regression analysis to determine trust amongst individuals. Tie is considered as a dependent variable being influenced by independent variables intimacy, intensity, friendship score and social distance. The results of the study indicate that tie is strongly influenced by intimacy and friendship score, while social distance and intensity have minimal influence. This trend offers new opportunities to build and predict strength among relationships. Between people it is apparent that microblogging services such as Twitter could become key applications in the sentiment analysis.

11. Future Scope

In this research work we have developed a linear as well nonlinear regression model and found that polynomial regression matched files the dataset of tweet score better as compared to the linear regression method .however, for future scope, we believe that it would be better if the dataset is also subjected to decision tree regression as it may happen that the data may come out to be more linear in nature in time of its degree

12. Acknowledgement

I am thankful to A P Nidhi, Assistant Professor, Swami Vivekanand Institute of Engineering and Technology, Banur, for providing constant guidance and encouragement for this research work.

Reference

[1] Twitter, http://en.wikipedia.org/wiki/Twitter
[2] Aditi Sharma Department of computer science and engineering SVIET PTU , Gurpinder Kaur Department of computer science and engineering SVIET PTU" *Understanding Social Ties in Context of Online Social Networks"*
[3] Tweets, https://dev.twitter.com/docs/platform-objects/tweets
[4] Users, https://dev.twitter.com/docs/platform-objects/users
[5] Naveen Sharma IIT Kharagpur MPI-SWS, Saptarshi Ghosh IIT Khargpur, MPI-SWS, Fabricio Benevenuto UFOP, Niloy Ganguly IIT Kharagpur *"Inferring Who-is-Who in the Twitter Social Network"*
[6] Cambridge Dictionaries Online [2009] *Definition Trust*, viewed on 22 Mar 2009. http://dictionary.cambridge.org/define.asp?key=85211&dict=CALD
[7] Jennifer Golbeck1 University of Maryland, College Park MD 20742, USA golbeck@cs.umd.edu WWW home page: http://mindswap.org" *Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering"*
[8] Eric Wang Department of Computing Macquarie University Sydney, Australia" *A Survey of Web-based Social Network Trust"*
[9] JENNIFER GOLBECK, JAMES HENDLER University of Maryland, College Park Computer Science Department College Park " *Inferring Trust Relationships in Web-based Social Networks"*