Trust metrics for online social media

Background :

The proliferation of Internet usage has resulted in huge amounts of information being made available online. A big contributor to this are social media networks such as Facebook, Twitter, Google +, Question and Answer platforms like Yahoo Answers, Quora and Stackoverflow which act as public boards of discussion. However, whilst the scale and variety of data has considerable value for commercial and social purposes, the quality, provenance, trust and validity is often questionable [1]. Opinions are frequently posted to social media sites and are often a mixture of fact, speculation or rumour whereas user-driven sites such as Wikipedia are often questioned for their trustworthiness. Blogs are used as sound boards for average users to disseminate information to their social circles and favourable opinion becomes fact which is in turn regarded as trustworthy information. Trust is also a primary factor in influence. Verified Twitter and Facebook profiles go some way to ensure that the identity of an entity is verified independently therefore ensuring the trustworthiness of that entity.

However, this does not go far enough in making sure that the information on social networks is truly accurate and therefore trustworthy. Trust can be quantified through the creation of algorithms and quality metrics which can be used to address the issues of validity and quality through the use of automated assessments. These metrics would be used to identify different aspects of an informational source and look for independent verification of the quality of the data. The metrics would need to be tested against real-world examples and vary depending on the platforms (e.g : Twitter, Facebook) and the situation/context. (e.g : Movie ratings might have different metrics and weights to a crisis situation); the nuances of language can change the meaning of a piece of information drastically through the use of slightly different words. An example of this would be “We anticipate the latest release to be announced shortly” versus “We expect the latest release to be announced shortly”. Although the words are synonyms, the first sentence regards the release as something that is probable. The second sentence regards the release to be more of a certainty.

The aim of this work is to provide a systematic method of verifying information on the internet independently and create a more trustworthy online experience. I understand that a uniform solution across all platforms would not be possible due to the complexities and variety of information shared across different platforms. However some key trustworthiness factors have been identified and I am planning to use machine learning techniques such as clustering, classification or regression to look for patterns and provide a reliable metric. There have been quite a few papers published in this area, some looking at very specific platforms, others adopting a more generalist approach.

Social media is now heavily relied upon as a source of information. This can vary from international breaking news to booking the next holiday. The common factor is that the information is often based on people’s experiences that are then shared online. Trustworthiness then becomes of paramount importance especially considering the fact that something that becomes popular on the internet is not necessarily true and vice versa. Hence my project will help improve upon current research and look at alternative solutions than those that are proposed and provide an implementation to test these out which I believe will help brands, people and the whole of the digital world to have a better online experience.

Open Questions :

1. How do we judge/measure the semantic completeness of the information we analyse?

2. What should be the level of importance given to factors affecting trustworthiness that are platform dependant / Context dependent (News Item, personal updates, public messages etc), or other.

Questions to study :

1. What are the factors that affect Trustworthiness in Online Social Media?
2. How do these factors relate to each other or affect each other, if there is a positive/negative correlation between the factors?
3. Come up with a suitable or reasonable metric and weight for each of these factors in predicting the trustworthiness of social media to a certain level of accuracy.
4. What should the formula / solution be? Would it be platform dependant or independent?
5. What is the best way of implementing this?
6. How would you then test it to verify your solution?

Proposed Method :

1. **Preparation :** Gather information required and other details about the project.

2. **Literature Review :**

a) From the reading I have done I have identified key factors that affect trustworthiness. Some of the papers that outline this are : "Information Quality and Trustworthiness : A topical state of the art review" published at the International conference on Computer Applications and Network security in 2011, "Believe it or not: Factors influencing credibility on the web" by C. Wathen and J. Burkell. Read through this data and select a few factors to research on or analyse their impact with respect to a couple of different social media platforms.

b) Look at existing papers that address this problem. There are some that are specific to certain social media sites. These will be helpful in getting an idea about how to tackle such problems.

c) Reading up on factors that have been analysed in depth and if any formulas have been suggested/proven better than others. Also check for successful experimentation and validation - all of which can help create a starting point for this project. This is will help to best chose the factors that I want to analyse.

3. Select a few of the metrics that we will focus on and do an in-depth study of each of them and how they affect trustworthiness. Currently I have chosen corroboration, popularity, verifiability and timeliness. The reason for this selection is that there are quite a large number of factors to study, which wouldn't be entirely possibly in four months and these four factors stand out as being slightly influential. However this is subject to change once more research has been done on this topic. Also these factors are, in comparison to other factors measurable and to some extent platform independent hence it is possible to come up with a reasonable metric and weights for each of these factors. The result of this is that I will have done an in-depth study on some factors which will be used in assessing trustworthiness. (Point #1 in section above)

4. After an in-depth study of the chosen factors we check to see if there are any correlations between these factors. We can do this by comparing results from the past or other methods which will be clearer when I actually start studying the factors. (Point #2 in section above)

5. From the study I would have done and from trial and error a reasonable metric can be attained. I would also analyse/understand how machine learning techniques have been applied to other factors and use some of the techniques i know to give weight to each factor. The details of this will only be clear after more reading / analysing. (Point #3 in section above)

6. Choose a social media platform (e.g Twitter) where the above factors play a significant role in deciding trustworthiness and develop a formula / solution that will let us determine trustworthiness with reasonable accuracy (>70%). (Point #4 in section above)

7. **Implementation** : The implementation of some proposed solutions would be in the Java Language however the specifics of the implementation haven't been thought through yet. This will be clearer when more research is done towards this project. (Point #5 in section above)

8. Manually test this against some samples - if it doesn't give results with accuracy of about 70% then refine the solution. (Point #6 in section above)

9. **Test** : Either use sample test data that is already available (e.g : Stanford Datasets) or pull the necessary data from the social site in question and test with the designed implementation. (Point #6 in section above)

Draft time table :

Date	April	May	June	July	August	September
Week 1		Reading / Research	Implementation / Research	Implementation	Report	Report
Week 2		Reading / Research	Implementation / Research	Implementation	Report	Report
Week 3	Reading / Research	Reading / Research	Implementation	Implementation	Report	
Week 4	Reading / Research	Start Implementation	Implementation	Implementation	Report	

Bibliography :

[1] Jason R. C. Nurse, Syed S. Rahman, Sadie Creese, Micheal Goldsmith, Koen Lamberts, "Information Quality and TrustWorthiness : A Topical State-of-the-Art Review," in 2011 International Conference on Computer Applications and Network Security (ICCANS 2011)

[2] Carlos Castillo, Marcelo Mendoza, Barbara Poblete, "Information Credibility on Twitter" in 20th International Conference on World Wide Web, 2011

[3] Joshua E. Blumenstock "Size Matters : Word Count as a Measure of Quality on Wikipedia," in 17th International Conference on World Wide Web, 2008

[4] Micheal P. O'Mahony, Barry Smyth , "Using Readability Tests to Predict Helpful Product Reviews," in Adaptivity, Personalization, Fusion Heterogeneous Information(2010)

[5] C. Wathen and J. Burkell, "Believe it or not: Factors influencing credibility on the web," Journal of the American Society for Information Science and Technology, vol. 53, no. 2, pp. 134–144, 2002.