

Determining Information Quality in Social Networks using Message Classification Techniques - an Exposé

Leon Edelmann

March 30, 2017

Social networks are playing an ever-growing role as an information source for a large public. The influence of said networks is spreading out to new demographics, becoming more approachable to a plethora of audiences of all ages and ranging from the tech savvy to technophobes. Unlike traditional news outlets however, social networks tend to offer anonymity to some extent to its users. This in turn, might eliminate the accountability, which traditional news outlets bear and is likely to be detrimental to the quality of information being spread.

The proposed study would examine such information exchanges in the form of messages (referred to hereinafter as *posts*) in social networks while trying to classify the information to a category as well as evaluating its reliability. The purpose is to present a novel method to classify information based on a range of features of the content itself as well as features of the author. The final product would be a classification of the post to a label {News; Not-News} and a measurement of the reliability of the information presented, either on a numeric scale or in binary form {Likely ; Unlikely}

The methodology of categorizing information to different labels corresponding to their usefulness was inspired by the work of Castillo, Mendoza and Poblete (2001) [1]. For the purpose of studying data-reliability, Text mining techniques as demonstrated by Go, Bhayani and Huang (2009)[2] are to be explored. The field of data quality in social networks has not yet been thoroughly explored and novel methods of quantifying such quality, if found, could prove to be very useful. Especially in light of the growing dependence of different interest groups on said networks. An interesting example of social network influence could be observed in the case when the official Twitter account of *The Associated Press* was hacked and a post saying that an attack on the white house occurred, and that president Obama was injured. Due to the automatization of many stock trading systems, the **S&P500** index underwent a crash causing about 130B \$ to be wiped. The capital was mostly restored, however the magnitude of such fluctuations is not to be understated [3]. This study would provide a survey of the existing state of research as well as examine, compare and experiment with implementing the different proposed techniques on sample data from social networks. A comparative survey in statistical terms, could then be compiled.

One possible experiment to study the credibility of information could be held in the form of an **electronic survey**. In the first stage, participants will fill out general information about themselves, such as age, gender, level of education, some measure for activeness in social networks and so on. A key-partition would be, erudition in the theme-subject, here being business and economics. The level of learnedness is to be differentiated relying on whether the said person, has had any experience academic, job-related or otherwise in the field. Afterwards, the participants would be presented with 10-15 posts consecutively. Each such post would be drawn at random from an existing pool. This pool would be compromised of posts extracted randomly from social media outlets, however being concentrated around a specific genre in this case business-related schemata. The participants would then be asked to classify each post to a category such as News, Chatter, Spam . Furthermore, each post should be rated in terms of **perceived credibility** on a numerical scale. Alternatively a corpus of posts could be manually categorized into labels creating a more subjective **training dataset**.

The results could then be studied to determine how this newly spread information is being received by members of social networks. Additionally, what factors make a post more/less credible. Possible characteristics of the post could be divided into categories such as: Characteristics-of-User, Characteristics-of-post, Characteristics-of-topic. Moreover, a further point-of-interest is whether proficiency in the field of economics is decisive in this regard. As a further stage the collected data is to be processed by machine learning algorithms, where ultimately an automatic classification mechanism would to be constructed.

As a case study, the micro-blogging platform **Twitter** would be used. The platform offers an API which allows tapping into the Twitters data stream and sending queries to their servers. These services are being offered free of charge to some extent. Namely, it is possible to tap to up-to 1% of the real-time stream of data, which flows through Twitter. These should be more than sufficient for the spectrum of this study. Alternatively, several corpora of previously collected Twitter-Stream-Data are also available on the internet. The bulk of research is to be conducted on a specific topic, mainly popular trends in the brach of E-Commrce platforms such as Amazon, eBay or Otto. With an additional aspect being, broadening the scope and using tapping to a non-thematized data stream from Twitter (the twitter Steaming API allows for a non-filtered query).

Another platform to be considered, depending on ease-of-access, would be **StockTwits**. The latter borrows the concept of Twitter and implements it on a narrower niche, namely the stock market and affiliated spheres. This platform also offers an API would could be used for data queries. Additionally, using StockTwits might prove useful when concentrating on business-related and commercial queries.

Moreover, an additional aspect to observe is the degree of correlation between social networks and the standing of companies. This includes but not restricted to: How influential is social media on the organization (measured in stock price), how credible is the information being spread, what characterizes users which influence the public view of the organization.

References

- [1] Carlos Castillo, Marcelo Mendoza and Barbara Poblete *Information Credibility on Twitter*. Proceedings of the 20th international conference on World wide web (WWW '11). ACM, New York, NY, USA, 675-684, 2011.
- [2] Go, A., Bhayani, R. and Huang, L *Twitter sentiment classification using distant supervision..* CS224N Project Report, Stanford, 1(12), 2009.
- [3] Christopher Matthews *How Does One Fake Tweet Cause a Stock Market Crash?*. <http://business.time.com/2013/04/24/how-does-one-fake-tweet-cause-a-stock-market-crash/>