Arabic Rumours Identification
By Measuring The Credibility
Of Arabic Tweet Content

Ahmad Yahya M. Floos, College of Computer and Information Sciences, King Saud University, Riyadh, Saudia Arabia

ABSTRACT

Twitter enjoys the fame of the most popular and widely used as a platform for socializing, including all aspects of life current affairs, religious ideas, political issues, scientific research, and general knowledge. Every single activity of day to day life and human behavior and values is lodged at this platform. Sending and receiving messages on Twitter (tweets) with is limited to 140 characters, In this research the author attempts to understand the characteristics of those Arabic rumour (falsified information stream) patterns. False tweets could be a rumour which is mostly recognized as a representative whose legitimacy, authenticity, precision and significance is either unverifiable or unreliable. Arabic rumours may propagate misinformation on social networks. In this research, the author illustrates the difficulty of Arabic rumour identification in twitter social platform by studying the impact based on Arabic tweet content. Furthermore, the author explains how these content features are too influential in measuring the credibility of those Arabic tweets.

KEYWORDS

Credibility, Machine Learning, Rumor, Social Network, Twitter

1. INTRODUCTION

This is evident that internet has become the most significant player for news in the last decade. This can be judged from the fact that only in United States of America, internet is the most popular source of news for the people under the age of thirty (30) while it is the second most popular medium for news after the television (TPR Center, 2008). Online Social Networks also called as OSNs like twitter and facebook are the recent innovative applications of this decade that influenced the humans largely. With the every passing day, the number of their users are growing. In 2011 only, the number of twitter users were more than 100 millions while the population of facebook users were more than 955 millions in 2012 (Securities and Exchange Commission, 2012). OSNs have totally change the way of communication among the people. People now can easily share their point of views and opinions to colleagues and friends and also spread the information to the community.

The popularity of social media like facebook and twitter has very greatly and deeply affected the journalism landscapes and news reporting. So, social media is not only used for the purpose of every day chat but also a very significant source of news and information sharing (Java, Song, Finin, and Tseng, 2007; Naaman, Boase, and Lai, 2010). Now, people refer towards social media as the major source of news (Laird, 2012). In breaking news situation, role of social media is even more prominent as the public crave for quick updated on the events and incidents in real time. Kwak et al. (2010) through his study showed that more than 85% topics on twitter are regarding news and current affairs. In addition, the accessibility, ubiquity, ease of use and speed of this social media has made it the invaluable and most important source of first hand information regarding news. In some cases,
twitter also proved to be very useful in disaster and emergency situations specifically for recovery and response (Vieweg, 2010).

In this work, we focused on measuring the credibility of Arabic News content published in Twitter. This study can be considered a first attempt to explore the credibility of Twitter content targeted for the Arabic language. For this purpose, we used the natural language processing and machine learning. Language is capability to talk, write and communicate and is one of the primary characteristic of human activities. As the learning of human languages progressed the perception of contact among machines was conceptualized, this served as the fundamental idea behind the natural language processing (NLP). The concept of natural language processing is to resolve and develop software that will examine, learn and produce natural human-languages. There are numerous software of natural language processing build through the long term. One of these softwares is Romurs Identifications, current users are searching through the major samples from every twitter microblogs (tweets) they read, people have no time to verify the authenticity of news. Given the extremely busy day to day lives and notherefore it be converted into a software for present-days.

There is two ways to perform text representation; indexing and term weighting. Indexing is the method of assigning indexing terms for a document, while terms weighting is the method of assigning weight to each term in the document. In this research, we use TF-IDF (Term frequency- inverse document frequency) to determine the weights for the terms contained in the document to determine the terms relevant the documents. We use documents that contain tweets for news and rumors, our job is to identify any new tweet to be defined as news or rumors according to the methods using tf-idf.

2. RELATED WORK

Related work has included the research related to the different roles of twitter in emergencies, researchers from the fields of network sciences about the propagation and diffusion of different information in social media networks, research from the fields of language processing about confining the propagation and diffusion of information in social media networks. We have also discussed different twitter credibility surveys, credibility of Arabic content and measurement of credibility of twitter information:

1. **Credibility Definition:** For the current study, the definition of credibility is taken from Reih (2010). Reih defines credibility as people’s assessment of whether information is trustworthy based on their own expertise and knowledge.

2. **Measurement of information Credibility:** In the assessment of credibility of information automatically, different approaches have been adopted by different researchers. For instance, machine learning supervised by learning methods, graphic analysis, statistical analysis and feature distribution have been potentially used to check the credibility. For the conveneience, the famous and most used techniques are presented in the Table 1.

3. **Credibility of Online News in Traditional Media and Blogs:** The In general, the users’ perception regarding the credibility of news published online appears to be positive. Usually, people trust the net as the major source of reading the news in comparrison with other media except newspapers (Flanagin and Metzger, 2000). This is fact now that internet has become the most significant player for news in the last decade. This can be judged from the fact that only in United States of America, internet is the most popular source of news for the people under the age of thirty (30) while it is the second most popular medium for news after the television (TPR Center, 2008). Online Social Networks also called as OSNs like twitter and facebook are the
Using Andragogy and Bloom’s Digital Taxonomy to Guide E-Portfolio and Web Portfolio Development in Undergraduate Courses

Mind the Gap
[www.igi-global.com/article/mind-gap/2781?camid=4v1a](www.igi-global.com/article/mind-gap/2781?camid=4v1a)