Analysis of Filipino Mood Swings within a Day using Tweets

Rodalyn A. Balajadia, Vincent Louie L. Maglambayan and Maria Teresa R. Pulido
Department of Physics, Mapúa University, Manila City, 1002, Philippines

Keywords: Data Analytics, Big Data Algorithm, Social Science and Implications for Big Data, APIs, User Evaluations and Case Studies, Microblogging, Twitter, Opinion Mining.

Abstract: We infer moods and how they vary with time using a dataset from the social media application Twitter. We used Python text mining techniques to gather all tweets originating from the Philippines within a span of 24 hours. From the dataset of around 130,000 tweets, we gathered the highest-frequency words and filtered out neutral words to come up with words that imply mood levels. We then plotted the density of keyword usage with respect to time, distinguishing between positive and negative moods. Our initial results of positive mood and negative mood trends are consistent with published studies regarding microblogging mood scales. The emergence of Big Data and the Internet of Things has greatly amplifed our ability not only to express ourselves but to understand each other.

1 INTRODUCTION

Moods affect physical and emotional well-being, creativity, decision-making, and immune response (Ashby, et al, 2002). Studies have shown that mood swings within the day heavily affect circadian rhythms of core body temperature (Boivin et al., 1997). In particular, positive moods have been found to peak twice: at noon and at evening (Hasler et al., 2008). Meanwhile, motivation is defined as a drive to behave in certain ways that comes from internal and external drivers and rewards such as mood (Deci and Ryan, 1985). Motivation is therefore intertwined with diurnal mood swings in correlation to human productivity.

Data mining enables us to make real-time measurements of global moods, such as social and political issues (O'Connor et al., 2010). In particular, researchers use the social networking service Twitter as "a global thought-stream on every topic imaginable" (Parr, 2009) with data gathering and analytical tools available to the public. Mislove et al. (2011) used Twitter comments to measure the "pulse of the nation" noting various time zones and days of the week. Meanwhile, Golder and Macy (2011) examined Twitter mood trends across diverse cultures accounting for global seasonal mood rhythms and individual-level diurnal moods. Bollen, Mao and Zeng (2011) used large-scale Twitter feeds to measure collective mood and predict the stock market. Self-reports of positive or negative moods may be conveyed through microblogging sites such as Facebook, Tumblr, or Twitter. Therefore individual behavior in real time can be collectively studied by examining semantic contents in such sources.

The purpose of this study is to analyze circadian moods of Filipinos using popular keywords with mood implications using Python text mining. This paper explored mood "pulses" confined within the coordinates of the Philippines by examining the density of popular keywords in tweets on an hourly basis. The gathering of tweets spanned for one weekday, or 24 hours.

Our study is limited only to publicly accessible accounts, in accordance with data privacy laws. Also, tweets are not representative of all age groups, as these are limited to retrospective self-reports of the younger generation (mostly millennials) and a low percentage of older people (Newberry, 2018). Lastly, this work will analyze word frequency only, looking at the literal meaning of tweets; figures of speech such as metaphors and sarcasm may be studied with more advanced language processing tools.

2 METHODOLOGY

We used a Python text mining program (Figure 1) to collect tweets in a streaming fashion via Twitter API.

Balajadia, R., Maglambayan, V. and Pulido, M.
Analysis of Filipino Mood Swings within a Day using Tweets.
DOI: 10.5220/0007749803640369
In *Proceedings of the 4th International Conference on Internet of Things, Big Data and Security (IoTBDS 2019)*, pages 364-369
ISBN: 978-989-758-369-8

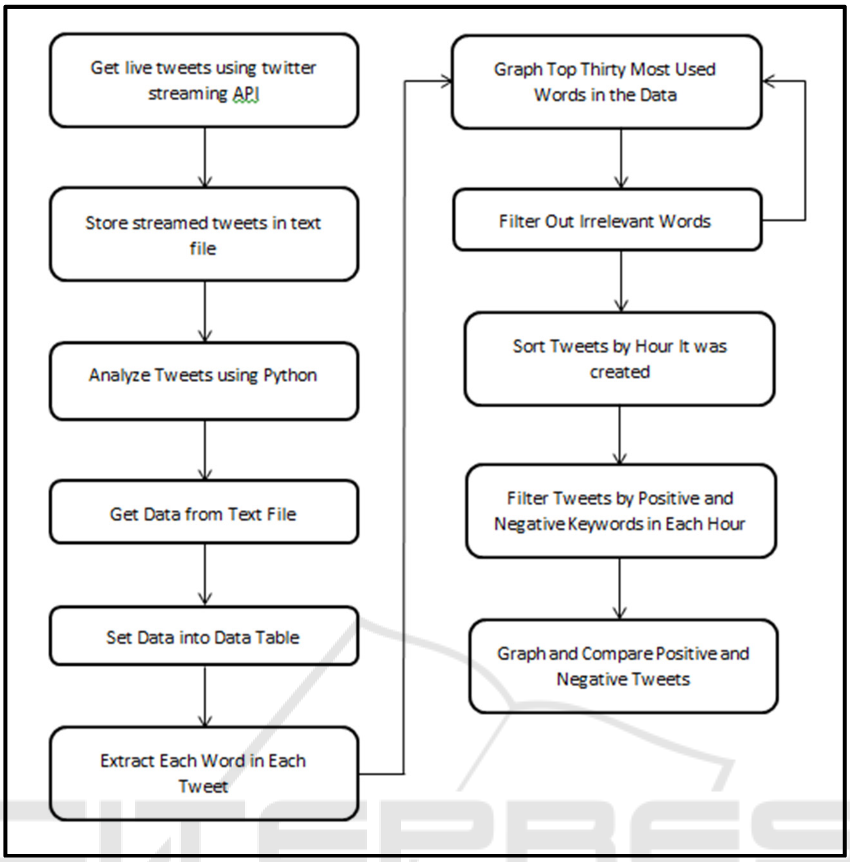

Figure 1: Flowchart of the program.

Figure 2: Sample tweet.

We gathered tweets posted within a particular 24-hour period, and only those originating from the Republic of the Philippines. Aside from the message itself, each tweet contains the handle or username of the sender, the date and location from where it was posted, the number of times it was retweeted and favorited, and other useful information (Figure 2).

We obtained the frequency of each word, or the number of times that a word appeared in the dataset. For the purposes of this study which focuses on mood swings, we manually filtered out fillers, conjunctions, and other emotionally neutral words.

We manually separated the resulting high-frequency words into two categories: those that correspond to positive mood (PM) and those that correspond to negative mood (NM). We plotted the frequencies of these PM and NM words with respect to time to look for possible trends corresponding to diurnal mood swings.

3 RESULTS AND DISCUSSION

We gathered approximately 130,000 tweets within a 24-hour period from the Philippines. Figure 3 shows the 30 words with the highest frequency. These include words evoking PM or NM, but many of the words are neutral or lack meaning.

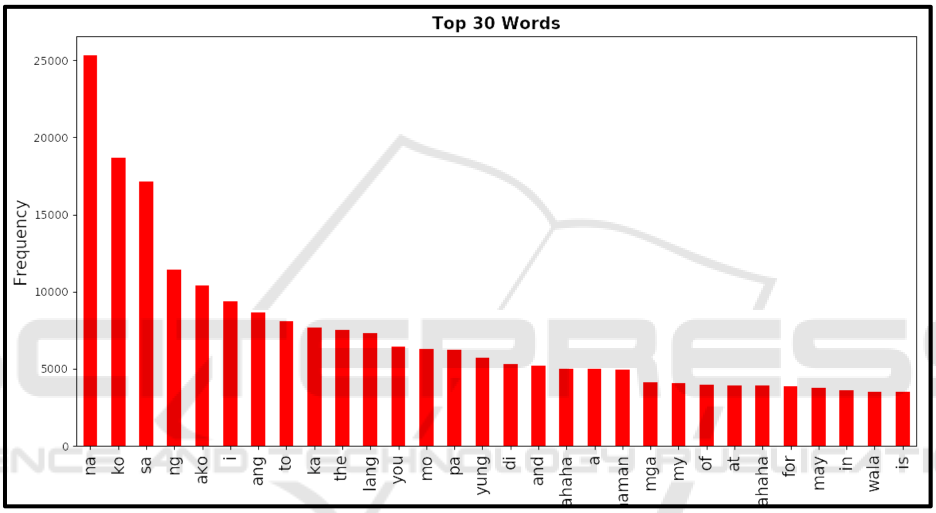

Figure 3: High-frequency words (unfiltered).

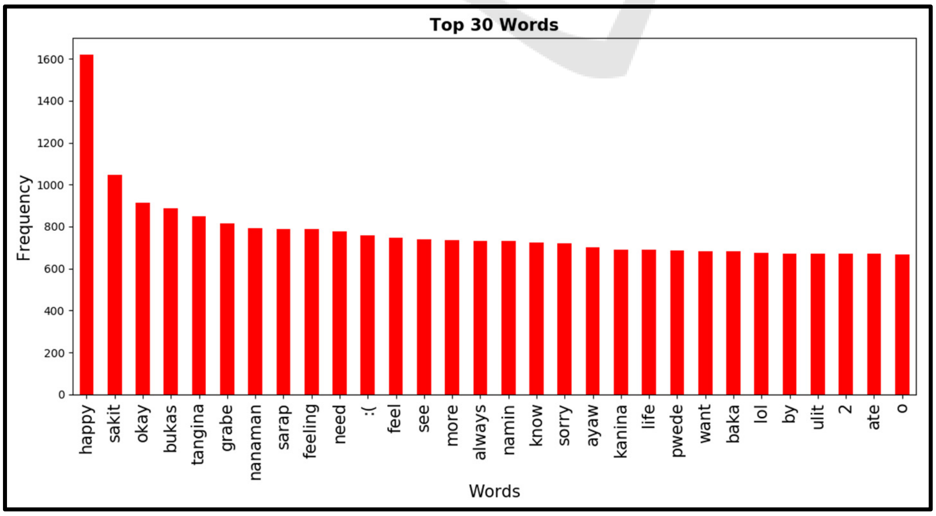

Figure 4: High-frequency words (filtered).

After filtering out neutral words, the 30 resulting high-frequency words became more evident in terms of the personal moods that they convey (Figure 4). We note that there were significantly more high-frequency words that evoked NM rather than PM.

The words were taken literally, so curse words like "*tangina*" were deemed negative while "*lol (laugh out loud)*" was deemed positive. Although "☹" is not exactly a word but an emoticon, we decided to include it as it also expresses a specific mood. There were still some neutral words such as "*okay*" and "*bukas* (tomorrow or open)" which were not considered as mood words. There were also words that conveyed deep emotion such as "grabe (very)", but cannot automatically be categorized as positive or negative.

Figure 5 shows the occurrence of the common PM words within the observation day. The number of PM tweets peaked during morning hours starting from 4 AM to 9 AM, where 9 AM had the highest peak. We note that the circadian sleep component was found to be elevated at waking and declines throughout the day (Boivin et al., 1997).

Meanwhile, PM tweets peaked at the usual lunch hour then declined during the afternoon (Figure 6). PM tweets gradually rose to two of its highest peaks during 7 PM and 9 PM, the common after-work hours, and had the steepest dip during post-midnight hours. In comparison, PM intuitively declined during work hours in a study by Golder and Macy (2011).

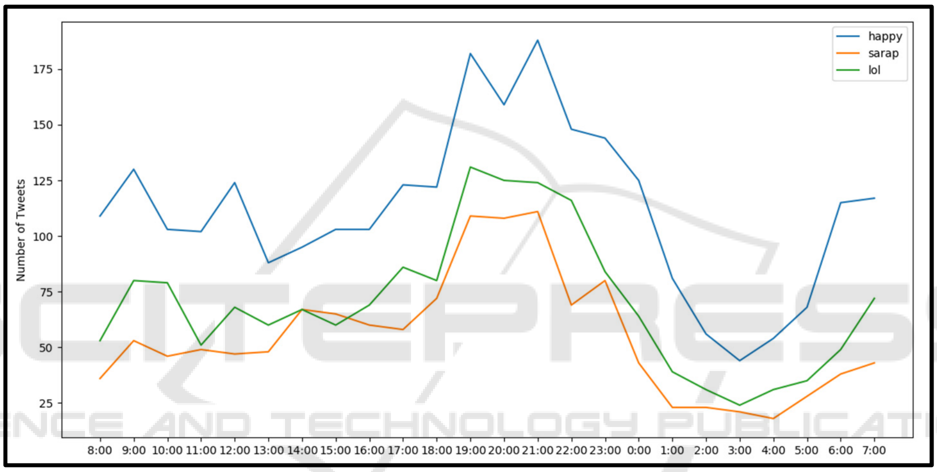

Figure 5: Positive Mood (PM) Time Scale.

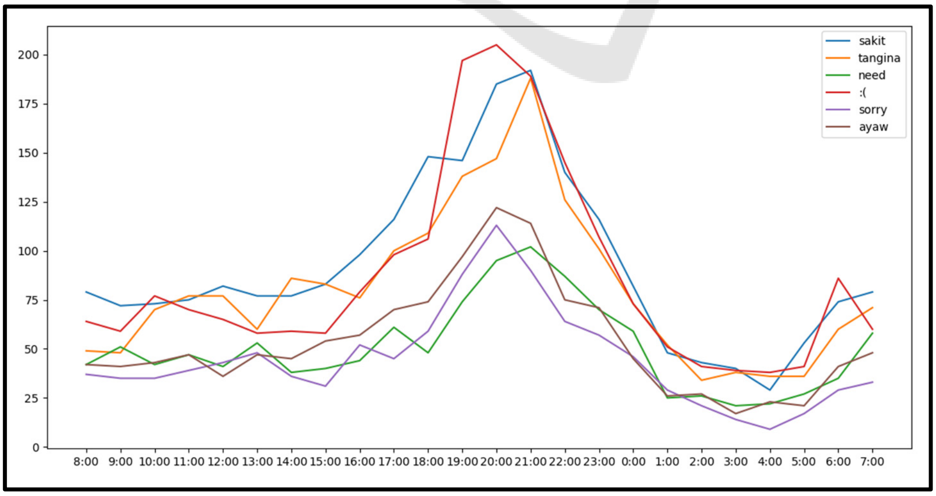

Figure 6: Negative Mood (NM) Time Scale.

NM words were a constant plateau during morning hours, gradually rising without relevant dips during the afternoon, until it reached its peak during after-work hours starting from 6 PM and peaking at its highest at 8 PM. A continuous decline was observed during sleeping hours, with a constant low plateau during post-midnight hours, and once again rising in the morning with a slight peak during 6 AM. This is relatively consistent with existing studies with their observation of the peaks and dips of NM (Golder and Macy, 2011).

Figure 7 combines the PM and NM words used in the previous two graphs. PM peaked during morning hours while NM had a relatively low plateau, but both consistently rose during after-work/after-school hours. PM peaks and dips were more abundant than the constantly rising and dipping scale of NM. Both had its steepest dips during post-midnight/sleeping hours, and both consistently rose as morning hours approached. Despite the abundance of NM in comparison to PM words, PM words were more frequent during morning hours. However, NM had its highest peak during after-work hours in comparison to PM.

We can delve deeper into this data by obtaining the high-frequency words that occur per hour, and then determine if they indicate positive or negative moods. Going further, we can determine the mood evoked by each tweet, which will give a much more accurate picture of mood and mood swings rather than individual words. We may also use sentiment analysis software to determine mood at a higher speed (especially for large datasets), and to ensure that the results are independent of possible bias from the researchers.

We note that the data was gathered on a Tuesday, which is commonly a work or school day. We can extend the study to investigate trends for other days of the week, or to consider user location, demographic traits such as age or gender, social network connections, and other related factors.

4 CONCLUSIONS

In conclusion, positive mood words exhibited erratic peaks and dips during morning hours and gradually declined during afternoon hours until it slowly rose to achieve its highest peak during after-work/after-school hours. Negative mood words consistently rose from morning until its highest peak during after-work hours without much erraticism. Despite negative mood words outnumbering positive mood words, positive mood words exhibited the highest frequency during morning hours. Both moods exhibited steep declines during post-midnight hours and consistently rose as morning approaches. Such trends found for tweets sent in the Philippines are consistent with previous studies involving Twitter users in other parts of the world.

We can expand this study by gathering data over a longer time period, or determine if there are regional differences in mood swings or at least in the words used to express them. We can also use sentiment analysis tools for more automated, objective measurement of positive and negative moods. The emergence of Big Data and the Internet of Things has greatly enhanced not only our ability to express ourselves but also the ability to understand each other.

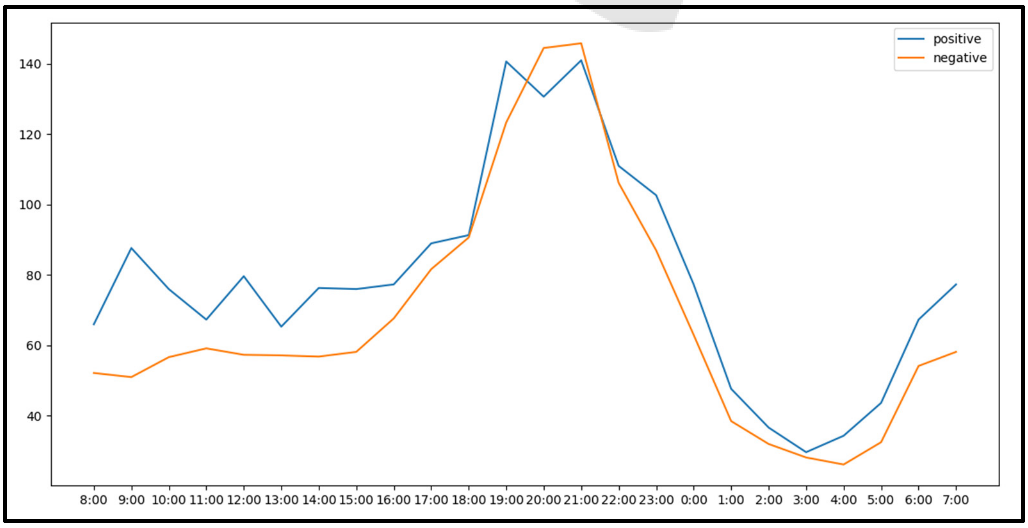

Figure 7: Comparison of PM and NM Time Scales.

ACKNOWLEDGEMENTS

We thank the Mapúa University Yuchengco Innovation Center for the resources in preparing this manuscript, and our colleagues and loved ones for their support. We also thank the organizers of the IoTBDS 2019 Conference for accepting this work and for the financial support.

REFERENCES

Ashby, F. G., Valentin, V. V. & Turken, A. U., 2002. Emotional cognition: From brain to behaviour.

Boivin, D. B., et al., 1997. Complex interaction of the sleep-wake cycle and circadian phase modulates mood in healthy subjects. *Archives of general psychiatry*, 54(2), pp.145-152.

Bollen, J., Mao, H. & Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of computational science*, 2(1), pp.1-8.

Deci, E., & Ryan, R. M., 1985. Intrinsic motivation and self-determination in human behavior. Springer Science & Business Media.

Golder, S. A. & Macy, M. W., 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051), pp.1878-1881.

Hasler, B. P., et al., 2008. Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality*, 42(6), pp.1537-1546.

Mislove, A., et al., 2011. Understanding the Demographics of Twitter Users. ICWSM, 11(5th), 25.

Newberry, C., 2018. 28 Twitter Statistics All Marketers Need to Know in 2018. [Online] Available from: https://blog.hootsuite.com/twitter-statistics/ [Accessed 30 December 2018].

O'Connor, B., et al., 2010. From tweets to polls: Linking text sentiment to public opinion time series. *Icwsm*, 11(122-129), 1-2.

Parr, B., 2009. 5 Terrific Twitter Research Tools. [Online] Available from: https://mashable.com/2009/05/03/twitter-research-tools/ [Accessed 03 January 2019].