Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

Describing and Classifying Post-Mortem Content on Social Media

Jialun "Aaron" Jiang, Jed R. Brubaker
Department of Information Science, University of Colorado Boulder
{aaron.jiang, jed.brubaker}@colorado.edu

Abstract

As the quantity of user profiles on social media grows, so does the number of post-mortem profiles. In this paper, we present a computational linguistic analysis of post-mortem social media profiles. Specifically, we provide an analysis of pre- vs. post-mortem language use, followed by a description of classifiers we developed that can accurately classify the mortality of social media profiles. These results shed initial lights into the ways in which people speak to the dead, and mark a first step toward accurately identifying mortality on a large scale.

Introduction

As the quantity of user profiles on social media grows, so does the number of post-mortem profiles. Without a means of identifying post-mortem profiles, data scientists and platform designers have limited ways of accounting for mortality in their work.

In this work, we provide a linguistic description of and introduce a machine learning-based identification method for post-mortem content on social media. Using 870,326 comments from 2,688 public profiles on MySpace as training data, we compare the performance of a series of supervised machine learning classifiers with three different feature sets, and show we are able to accurately classify post-mortem user profiles.

Our results mark a first step toward automatic identification of post-mortem profiles on social media. Automatic identification also suggests new ways to identify post-mortem profiles, enabling data scientists and platform designers to account for mortality in their datasets.

The Language of Death and Bereavement

Previous research has studied language use around death using computational methods in a variety of contexts, from classifying reports of death (Imane and Mohamed 2017) to identifying suicidal signals (Mulholland and Quinn 2013; Desmet and Hoste 2013, 2014; Pestian et al. 2010; Huang et al. 2017), and achieved promising results. Research of bereaved language has also revealed important patterns of sentiment and linguistic style.

Researchers have found success using computational linguistics tools like "Linguistic Inquiry Word Count" (LIWC) [1], a common language analysis package that provides dictionaries for parts of speech and punctuation, as well as psychosocial and social processes. Brubaker et al. (2012) found that in comments made to post-mortem social media profiles, words from LIWC's first person singular pronoun, past tense verb, adverb, preposition, conjunction, and negation categories appeared more frequently in emotion distressed comments; in terms of sentiment, distressed comments also showed higher use of anger words. These results suggest highly emotional bereaved language is accompanied by distinctive linguistic signals. Likewise, in Ma et al.'s (2017) study of writing on an online health community, the they were able to use journals (blog-like entries that included obituaries and funeral announcements) alongside bereaved content to successfully classify mortality using words from LIWC's death dictionary (e.g., died, funeral, grave). Together, these studies show great potential in using computational linguistics to analyze distinctive changes in language after death, and serve as inspiration for our work classifying the mortality of social media content.

[1] http://liwc.wpengine.com/how-it-works/

Data

The data used for the work we present in this paper are a subset of comments collected from profiles of 13,200 deceased MySpace users in April 2010. Deceased profiles were identified using MyDeathSpace (MDS), a website dedicated to connecting obituaries and/or news of deaths to existing MySpace profiles. At the time of collection, MDS contained more than 15,000 user-submitted entries, as well as notes including additional information and links to other online content (e.g., newspaper articles).

For this study, we first limited our sample to comments posted to publicly visible profiles of users who lived in the United States and who had been dead for at least three years. The subset used for this study consists of 1,463,096 randomly selected comments posted to 4,725 publicly visible profiles by 303,276 unique comment authors.

To verify mortality, each profile was hand-checked by one of the researchers using comments posted to the profile or changes made to the profile itself (e.g., adding "RIP" or "In memory of" to the profile name). Further data preprocessing reduced the number of comments and consequently the number of associated profiles in our training data, which we describe next.

During pre-processing, we removed comments identified as spam from our dataset. We aggressively removed spam by removing all comments without English characters or contained links to other websites.

Next, we cleaned the remaining data by removing all HTML tags, all names, all nicknames, and all non-English characters from the comments. Names were identified using a list of first names published by US Census in 1990, and a list of common nicknames (e.g., Ben for Benjamin)[2]. We then substituted common Internet word abbreviations with their full forms in all comments (e.g., "luv" was substituted by "love"). We also substituted different variants of the phrase "rest in peace" with the abbreviation "rip."

Finally, we removed profiles that had less than five pre-mortem or post-mortem comments to ensure each profile has enough data to be classified, which further reduced the number of comments and profiles. Our spam cleaning process along with extra pre-processing excluded 592,770 comments, leaving us 870,326 comments for the final dataset. Descriptive statistics of the final dataset used for our classification tasks are shown in Table 1.

Total comments	870,326
Total profiles	2,688
Post-mortem comments	324,089 (37.24%)
Pre-mortem comments	546,327 (62.76%)
Average comments per profile	323.78
Median comments per profile	169.50
Average words per comment	32.74
Average post-mortem comments per profile	120.57
Average pre-mortem comments per profile	203.21

Table 1: Descriptive statistics of the final dataset.

Analysis of Post-Mortem Language

We started by examining differences between the length of pre- and post-mortem content. The average word count of post-mortem comments ($\mu = 48.79$) is significantly greater than pre-mortem comments ($\mu = 23.22$) (Mann-Whitney U $= 5.9 \times 10^{10}$, $p < 0.001$), confirming what Getty et al. (2011) found on a smaller dataset.

Next, we looked at differences in linguistic characteristics. Existing work shows that linguistic characteristics are strong indicators of bereaved communication (Getty et al. 2011) and emotional distress (Brubaker et al. 2012) on post-mortem profiles. Following this work, we used LIWC to provide a score between 0-100 for each body of text (normalized to 0-1 in our analysis), indicating the proportion of words in the text contained in the dictionary for each given category.

We conducted Mann-Whitney U-tests to compare the mean differences in the LIWC metrics studied in Getty et al. (2011) and Brubaker et al. (2012). To account for multiple comparisons, p-values were corrected using the Holm-Bonferroni method. Descriptive statistics, p-values, and effect sizes (Cohen's d) are shown in Table 2.

These results show that the linguistic characteristics identified as important in previous work are also strong indicators of mortality in our dataset. Post-mortem comments showed higher word count, greater use of second-person pronouns, and sadness words with medium effect sizes ($|d| > 0.5$), and greater use of first-person pronouns, negative emotion, and present tense with small effect sizes ($|d| > 0.2$). Higher word count may indicate that post-mortem comments take more time to construct, and higher use of present tense, first-person pronouns may indicate post-mortem comment authors were writing with low psychological distance (Cohn, Mehl, and Pennebaker 2004) as would be expected if they were personally connected to the deceased, which is also indicated by higher use of second-person pronouns. Finally, higher use of negative emotion, sadness words may, understandably, indicate that comment authors were expressing grief.

[2] https://deron.meranda.us/data/

LIWC Metric	Pre-mortem	Post-mortem	p	d
Word count	0.232	0.488	*	-0.523
First person pronoun	0.067	0.092	*	-0.361
Second person pronoun	0.051	0.093	*	-0.549
Past tense	0.028	0.034	*	-0.122
Adverb	0.056	0.058	*	-0.023
Preposition	0.078	0.075	*	0.041
Conjunction	0.044	0.048	*	-0.090
Negation	0.014	0.012	*	0.045
Social processes	0.140	0.191	*	-0.389
Positive emotion	0.079	0.067	*	0.129
Anger	0.011	0.005	*	0.165
Sadness	0.007	0.040	*	-0.682
Negative emotion	0.026	0.050	*	-0.352
Articles	0.029	0.026	*	0.068
Six-letter words	0.089	0.079	*	0.089
Discrepancy	0.014	0.015	*	-0.028
Present tense	0.104	0.124	*	-0.215

*Table 2: Mann-Whitney U-tests results and descriptive statistics of pre- and post-mortem comments. * $p < 0.0001$*

Classifying Profiles

Having confirmed significant differences between pre- and post-mortem language in our dataset, we proceeded to develop classifiers. In this section, we present a machine learning classification task to classify if a profile is pre-mortem or post-mortem (i.e., whether the profile owner is alive or dead). We discuss the classification methods used in the task and their associated classification results.

Classification Task

We sought to classify whether a profile is pre-mortem or post-mortem based on all comments posted on the profile. Our unit of analysis here is the concatenated comments of each profile. Since all the profiles in our dataset are post-mortem, we randomly selected half of the profiles to be "pre-mortem" and excluded any comments added post-mortem. The remaining half became the post-mortem profiles. A naïve but understandable assumption might be that "rip" unambiguously indicates someone's death. Therefore, we first developed a simple baseline rule-based classifier: Classify a profile as post-mortem if the text from the profile contains the word "rip", and pre-mortem otherwise. To compare with the baseline classifier, we then implemented four commonly used text classifiers: Multinomial Naive Bayes (NB), Logistic Regression (LR), Linear SVM (SVM), and Boosted Trees. NB, LR, and SVM were implemented with the Python machine learning library scikit-learn; Boosted Trees was implemented with the XGBoost system developed by Chen & Guestrin (2016).

For these four classifiers, we compared three different feature sets:

(1) n-gram features (n = 1,2,3) with TF-IDF weights;
(2) features derived from computational linguistic tools (CLT): style, topic, and sentiment measures from LIWC (Tausczik and Pennebaker 2010), Empath's default categories (Fast, Chen, and Bernstein 2016) and VADER (Hutto and Gilbert 2014); and
(3) the combination of (1) and (2).

We removed VADER's *compound* metric from our feature sets for two reasons: First, its value can be negative, which is incompatible with our chi-squared feature selection and Naive Bayes classifier. Second, the *compound* metric's collinearity with VADER's *pos*, *neg*, and *neu* also makes it a redundant feature.

For feature sets (1) and (3), we additionally conducted chi-squared feature selection due to the large number of features derived from our n-gram model. The number of features that we used in feature selection ranged from 1,000 to 10,000 in 1,000 intervals. We evaluated our classifiers using F1 scores and 10-fold cross validation.

Table 3 shows the best performance of each classifier from feature selection in each feature set. The baseline classifier performed with an F1 score of 0.74, but still not as

		Accuracy	*F1*	*Precision*	*Recall*
	Baseline	0.696	0.736	0.697	0.778
n-gram	NB	0.835	0.837	0.901	0.782
	LR	0.856	0.858	0.946	0.785
	SVM	0.862	0.866	0.952	0.794
	XGBoost	0.876	0.881	0.942	0.827
CLT	NB	0.593	0.720	0.578	0.953
	LR	0.750	0.769	0.775	0.764
	SVM	0.789	0.793	0.846	0.747
	XGBoost	0.821	0.828	0.884	0.779
n-gram + CLT	NB	0.846	0.840	0.961	0.746
	LR	0.856	0.858	0.939	0.790
	SVM	0.865	0.865	0.952	0.793
	XGBoost	**0.874**	**0.881**	**0.940**	**0.829**

Table 3: Classification metrics of profile level classifiers. Classifier with the best F1 score is shown in bold.

well as the other classifiers. Our baseline classifier achieved a recall of 0.78, in other words, there was at least one "rip" posted to 78% of the post-mortem profiles, which is consistent with the norm of how people speak to the deceased offline. Overall, XGBoost with 4,000 features from feature set (3)—n-grams and linguistic tools features combined—had the best performance, with an F1 score of 0.881.

Discussion & Future Work

The results of the classification tasks show that post-mortem comment messages and profiles can be identified with a high degree of accuracy. In our classification task, the feature set that combined n-grams and linguistic measures had the best performance. This result indicates that both specific words and linguistic style are important for identifying post-mortem status.

While we are able to classify the mortality of a profile, it is also important to be able to do so more granularly and quickly so that platforms can provide better support for the survivors. Therefore, our future work includes classification on the comment-level, as well as accurate classification with minimal amount of post-mortem content.

One limitation of this work may be the generalizability of our classifiers to other contexts. Our data were drawn from MySpace in 2010. Changes in linguistic practices since then may constrain our findings. Given this, applying classifiers such as ours to datasets from other platforms, as well as examining how post-mortem content has changed (or not) over time, are also clear steps for future work.

References

Brubaker, J.R., Kivran-Swaine, F., Taber, L., and Hayes, G.R. 2012. "Grief-Stricken in a Crowd: The Language of Bereavement and Distress in Social Media." In *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, edited by John G. Breslin, Nicole B. Ellison, James G. Shanahan, and Zeynep Tufekci, 42–49. Dublin, Ireland: The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4622/4965.

Chen, T., and Guestrin, C. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. KDD '16. New York, NY, USA: ACM. doi:10.1145/2939672.2939785.

Cohn, M.A., Mehl, M.R., and Pennebaker, J.W. 2004. "Linguistic Markers of Psychological Change Surrounding September 11, 2001." *Psychological Science* 15 (10): 687–93. doi:10.1111/j.0956-7976.2004.00741.x.

Desmet, B., and Hoste, V. 2013. "Emotion Detection in Suicide Notes." *Expert Systems with Applications* 40 (16): 6351–58. doi:10.1016/j.eswa.2013.05.050.

Desmet, B., and Hoste, V. 2014. "Recognising Suicidal Messages in Dutch Social Media." In *LREC 2014 - NINTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION*, 830–35. http://hdl.handle.net/1854/LU-6993527.

Fast, E., Chen, B., and Bernstein, M.S. 2016. "Empath: Understanding Topic Signals in Large-Scale Text." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4647–4657. CHI '16. New York, NY, USA: ACM. doi:10.1145/2858036.2858535.

Getty, E., Cobb, J., Gabeler, M., Nelson, C., Weng, E., and Hancock, J. 2011. "I Said Your Name in an Empty Room: Grieving and Continuing Bonds on Facebook." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 997–1000. CHI '11. New York, NY, USA: ACM. doi:10.1145/1978942.1979091.

Huang, X., Xing, L., Brubaker, J.R., and Paul, M.J. 2017. "Exploring Timelines of Confirmed Suicide Incidents Through Social Media." In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 470–477. Park City, UT: IEEE. doi:10.1109/ICHI.2017.47.

Hutto, C.J., and Gilbert, E. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." In *Eighth International AAAI Conference on Weblogs and Social Media.* http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109.

Imane, A., and Mohamed, B.A. 2017. "Multi-Label Categorization of French Death Certificates Using NLP and Machine Learning." In *Proceedings of the 2Nd International Conference on Big Data, Cloud and Applications*, 29:1–29:4. BDCA'17. New York, NY, USA: ACM. doi:10.1145/3090354.3090384.

Ma, H., Smith, C.E., He, L., Narayanan, S., Giaquinto, R.A., Evans, R., Hanson, L., and Yarosh, S. 2017. "Write for Life: Persisting in Online Health Communities Through Expressive Writing and Social Support." *Proc. ACM Hum.-Comput. Interact.* 1 (CSCW): 73:1–73:24. doi:10.1145/3134708.

Mulholland, M., and Quinn, J. 2013. "Suicidal Tendencies: The Automatic Classification of Suicidal and Non-Suicidal Lyricists Using NLP." *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 680–84.

Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., and Leenaars, A. 2010. "Suicide Note Classification Using Natural Language Processing: A Content Analysis." *Biomedical Informatics Insights* 2010 (3): 19–28.

Tausczik, Y.R., and Pennebaker, J.W. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology* 29 (1): 24–54. doi:10.1177/0261927X09351676.