The Workshops of the Thirty-Second
AAAI Conference on Artificial Intelligence

Detecting Personal Experience Tweets for Health Surveillance Using Unsupervised Feature Learning and Recurrent Neural Networks

Shichao Feng,[1,2] Keyuan Jiang,[2] Jiyun Li,[1] Ricardo A. Calix,[2] Matrika Gupta[2]
[1]Computer Science and Technology, Donghua University, Shanghai 201620, China
[2]Computer Information Technology and Graphics, Purdue University Northwest,Hammond, Indiana 46323, USA
uranusfeng1992@gmail.com, kjiang@pnw.edu, jyli@dhu.edu.cn, rcalix@pnw.edu, gupta.matrika@gmail.com

Abstract

Given its easy accessibility and prevalence, Twitter has been actively used as an alternative data source for health surveillance research, and personal health experiences play an important role in such surveillance activities. Therefore, there is a need to develop efficient and effective methods to identify Twitter posts related to personal health experiences. In this work, we present a method which combines word embeddings, convolutional, and Long Short-Term Memory (LSTM) recurrent neural networks to detect personal health experience tweets. The word embedding and convolutional layers serve as a pre-processing step for unsupervised feature learning. This step helps to eliminate the need for feature engineering. We studied three distributed word representation methods: word2vec, fastText, and WordRank to represent the tweet texts in a vector space model. Vectors of the word representations were later used in a convolution layer for further pre-processing, and were fed to an LSTM based Recurrent Neural Network (RNN) model for classification. Our results showed that approach outperforms, with a significant margin, conventional classifiers that used human engineered features. The RNN based model had a significant improvement in precision compared to the other methods (by 123%). This improvement helps to detect more true positive Personal Health Experience tweets.

Introduction

The World Health Organization defines that public health surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice (WHO). Information obtained from the population is most valuable and important to surveillance activities. Information directly reported by the population is of significant importance in understanding their health related issues, and there is a need to develop efficient ways of analyzing this data.

Emergence and prevalence of social media such as FaceBook and Twitter have made it possible for people to share their personal experiences on social media. Studies have shown that general purpose social media such as Twitter can be used for surveillance of health-related issues (Dredze 2012) and public health surveillance (Yepes, MacKinlay, and Han 2015). There have been a number of studies that validate the use of Twitter for surveillance of health related issues. Examples include: influenza pandemics (Chew and Eysenbach 2010; Signorini, Segre, and Polgreen 2011; Collier, Son, and Nguyen 2011; Bilge et al. 2012; Nagel et al. 2013; Gesualdo et al. 2013; Broniatowski, Paul, and Dredze 2013; Fung et al. 2013; Nagar et al. 2014), Haitian cholera outbreak (Chunara, Andrews, and Brownstein 2012), Ebola outbreak (Odlum and Yoon 2015), non-medical use of a psychostimulant drug (Adderall) (Hanson et al. 2013), drug abuse (Chary et al. 2013), smoking (Sofean and Smith 2012), suicide risks (Jashinsky et al. 2014), migraine headaches (Nascimento et al. 2014), pharmaceutical product safety (Freifeld et al. 2014; Coloma et al. 2015; Jiang and Zheng 2013; Sarker et al. 2015), disease outbreaks during festivals (Yom-Tov et al. 2014), detection of Schizophrenia (McManus et al. 2015), foodborne illness (Harris et al. 2014), dietary supplements side effects (Jiang et al. 2017), and even dental pains (Heaivilin et al. 2011). Many of these health surveillance studies involve using the information reported by patients who shared their personal health experiences on social media.

Personal health experiences are of unique importance in health surveillance activities, because they provide the first-hand encounters of changes in health conditions. However, it is challenging to extract personal experiences from the vast amount of social media posts. General purpose social media data are known for their noisiness, and the data gathered can have significant amount of posts irrelevant to the health issues to be monitored, and unrelated to any personal health experiences. A common challenge identified in health surveillance studies using social media is the difficulty in separating the useful or “on-topic” posts from the majority of the irrelevant posts.

Processing and analyzing Twitter data with natural language processing (NLP) and machine-learning algorithms pose unique challenges. As a micro-blogging service, Twitter limits each post to 140 characters, making it difficult to extract features from the text. Besides, Twitter users do not follow the grammatical and spelling rules, rendering poor performance in analyzing the tokens and semantics. Therefore, many conventional machine-learning methods are inadequate if applied to Twitter data. With the flexibility offered to Twitter textual data, significant efforts and human intelligence are needed to identify useful and important features for classification. It is observed that the conventional methods are based on statistics of features without considering the semantics of tweet textual data, leading to low accuracy in classification.

Traditionally, feature extraction has been a laborious task performed by domain knowledge experts usually referred to as feature engineering. In the past, these methods have involved lexical or syntactic approaches that extract frequency based statistics about tokens in a text. Extracting the higher level semantics from the text has been a much more challenging issue that can be very labor intensive. As a result, the difficulty in extracting semantic data can lead to low accuracy in classification.

In recent years, there has been a shift from feature engineering to more automated ways of extracting features; especially, in the text processing domain. In particular, distributed vector representations (or word embeddings) are useful to perform the feature extraction from text.

The methodology presented in this work uses word embeddings to perform feature extraction from twitter texts. The word embeddings are further processed using a convolutional neural network layer which helps to compress the data and to help discover feature representations. After the convolutional layer, the resulting vector representations per tweet are used for supervised learning classification using a Long Short-Term Memory (LSTM) recurrent neural network. In this study, we tested three different word embedding models (word2vec, fastText, and WordRank) to represent the words in the tweets.

The results of our experiments show that our methodology, using a deep neural network with convolutional and LSTM-RNN layers, performs better than conventional methods. The methodology achieves a 123% increase in precision and significant improvements in all other performance metrics.

Method

In this work we propose the use of a deep neural network to perform classification and detection of personal health experience tweets. We use an annotated data set of tweets to train and test the model. The deep neural network consists of both convolutional and recurrent neural network layers. The use of a recurrent neural network layer, in particular, is important because it helps to capture the sequential semantics of the words in each tweet. The convolutional layer serves as a preprocessing step to the recurrent neural network layer which helps to further compress the inputs and to possibly learn additional salient features in each text. The data processing pipeline is illustrated in Figure 1.

The inputs to the deep neural network are vectors representing each tweet. Word embedding methods were used to convert the tweet texts into dense vector representations. In this case, each word in the text is converted into a 200 dimensional dense vector. The word embedding step requires the construction of a vector space where every word in a corpus can be represented. These vector spaces are created based on word co-occurrence in the corpus using efficient neural network methods such as word2vec (Mikolov et al. 2013).

An unlabeled corpus of 22 million tweets was used to create the word vector space. This 22 million tweet corpus was constructed by retrieving medicine-related tweets from August 25, 2015 through December 7, 2016 (using the Twitter Streaming APIs). It is important to note that this word embedding corpus is not the same corpus used to train and test the model. This corpus of 22 million tweets is unlabeled and was used for the word embedding step only.

Word embedding approaches benefit from using large amounts of data (Mikolov et al. 2013) and, as such, this was the rationale for collecting the 22 million medicine related tweets. Given that twitter data is very noisy, we should note that some pre-processing of this unlabeled data set was needed to arrive at the 22 million tweets used. Pre-processing included: removing re-tweets, non-English tweets, URLs, and punctuations.

Each tweet was converted into its vector space representation using the unlabeled word vector space. In this case, each tweet can be thought of as a variable length sequence of 200 dimensional vectors where each word is a vector (of size 200). These input vector sequences are then passed to the convolutional layer of our deep neural network. Convolutional network are usually used in the context of images of fixed or consistent size. To address this issue, we decided to treat each tweet as a matrix (image) of 48x200 dimensions. In this case, the number 48 represents the number of words that we allow per tweet and 200 is the vector size of each word. We use padding for tweets with less than 48 words and truncation for tweets with more than 48 words.

Personal Experience Tweets

For the purpose of this work, a personal experience tweet (PET) is defined as a tweet that describes a persons encounters, observations, and events related to his or her life (Jiang, Calix, and Gupta 2016). The PETs are important for health surveillance because they can reflect the changes in a persons physical conditions. These can include experiences with an illness, a disease or a medical treatment. The following are examples of personal experience tweets that address health related issues.

*"accidentally took way too much **ibuprofen** for my cold & now i'm at work drowsy..... best of luck to me"*

*"my headache going away. this **ibuprofen** really worked for me"*

*"**Ibuprofen** did the trick and my shoulder is back to normal this morning. Just one of those stupid 'I am ancient' things, I guess."*

Word embedding

To investigate the importance of the word embedding techniques, we performed the analysis using 3 well known word embedding techniques. The three methods used were: word2vec, fastText and WordRank.

Word embeddings have become an active area of research in recent years and are widely used now for natural lan-

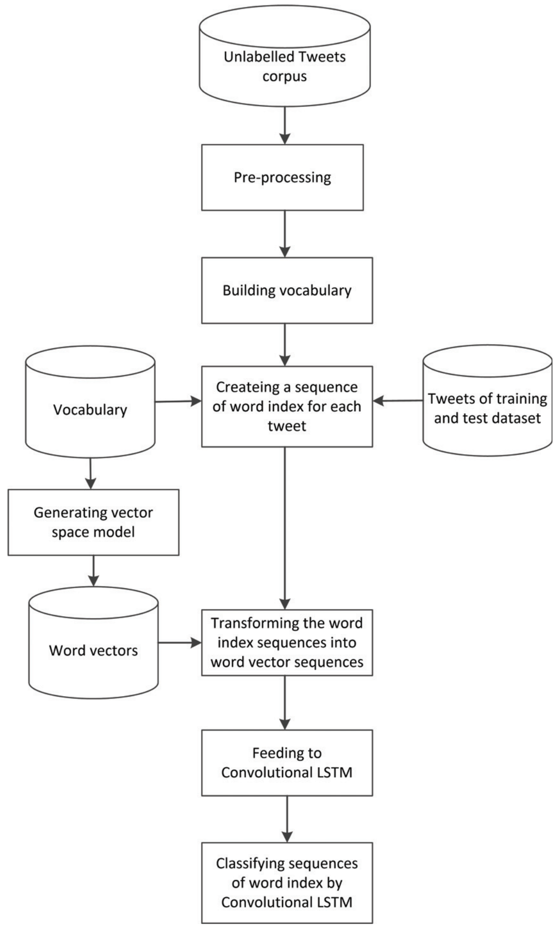

Figure 1: Flow chart of identifying the personal experience tweets

guage processing. In general they are preferred over human based feature engineering tasks because they are less time consuming and can achieve good results. Word embedding techniques have been used in many fields such as in generic text classification (Lev, Klein, and Wolf 2015), and Twitter sentiment analysis (Severyn and Moschitti 2015). The quality of the word embedding method can influence the performance of the classifier. Therefore, to address this issue we have run the proposed method using three word embedding techniques which are word2vec, fastText, and WordRank .

Word2vec (Mikolov et al. 2013) and fastText (Bojanowski et al. 2016) are similar embedding approaches. The main difference is that word2vec treats each word as an atomic unit whereas fastText treats each word as a composite of smaller atomic units such as character n-grams. Each vector in fastText is therefore a sum of the individual character n-gram vectors. Intuitively, this can be helpful for rare words that consist of several meaningful units. For example, the word “superstructure”, which is a compound word, may not be in the vocabulary. In word2vec, this word may be missed. With fastText, some semantics of this word may be recovered if the word is represented by its smaller atomic units that may be present in other words in the vocabulary such as “struc” and “ture”, for instance. Finally, WordRank sets up the word embedding optimization as a ranking problem (Ji et al. 2015). That is, given a word “w”, WordRank tries to predict an ordered list of words c1, c2, , cn where the words are ordered by frequency of co-occurrence with “w”. One advantage of WordRank is that it can achieve similar results to the other methods while using fewer examples.

Convolutional Long Short-term Memory Classifier

Figure 2 presents the architecture of the proposed deep neural network. As previously indicated the proposed deep neural network consists of a convolutional layer for feature discovery and an RNN layer for sequential classification. RNNs have been used extensively in natural language processing tasks such as in (Mikolov et al. 2010; Lai et al. 2015) and can be traced back to early work such as (Elman 1990). Essentially, they are sequential based techniques that look at the current input to the model plus the results from the processing of previous inputs.

The most common and best performing RNN approach is the long short-term memory (LSTM) based RNN network (Gers, Schmidhuber, and Cummins 1999). LSTM networks perform better than RNN networks because they are RNN networks with more optimized components (LSTM units). In essence they have LSTM units that can better retain previous information (without modifying it during the training process). They have been proven to perform really well in many studies.

The convolutional layer used in our model was used to discover better representations of the features. Convolutional neural networks have been used in text processing task such as character-level feature extraction (Zhang, Zhao, and LeCun 2015), and sentences classification (Wu, Yin, and Liu 2017).

As previously indicated, in our convolutional LSTM model, we transformed each tweet into matrices of 48x200 dimensions. Tweets with less than 48 words were padded by an all 0s vector and unknown tokens used an all 1s vector. We use 1 CNN layer with slide window of size 5 and 1 maxpool layer. After convolution, the tweets are represented as matrices of size 44x128.

The analysis was performed using Keras[1] and genism[2]. Keras is a front-end library which contains Googles TensorFlow[3] interface, and Genism is a free python package which contains the word2vec model and the interface to implement the fastText and WordRank model. In the LSTM model, we used L2 regularization and class-weight setting to adjust the parameters, and we trained the model for over 200 epochs to observe which epoch had the best performance. Finally, we chose 4 epochs to train the model.

Datasets

Vector space models were constructed from a corpus of 22 million tweets. To train LSTM model and test the model, 12,331 tweets were randomly selected from the corpus and

[1]https://keras.io
[2]https://radimrehurek.com/gensim
[3]https://www.tensorflow.org

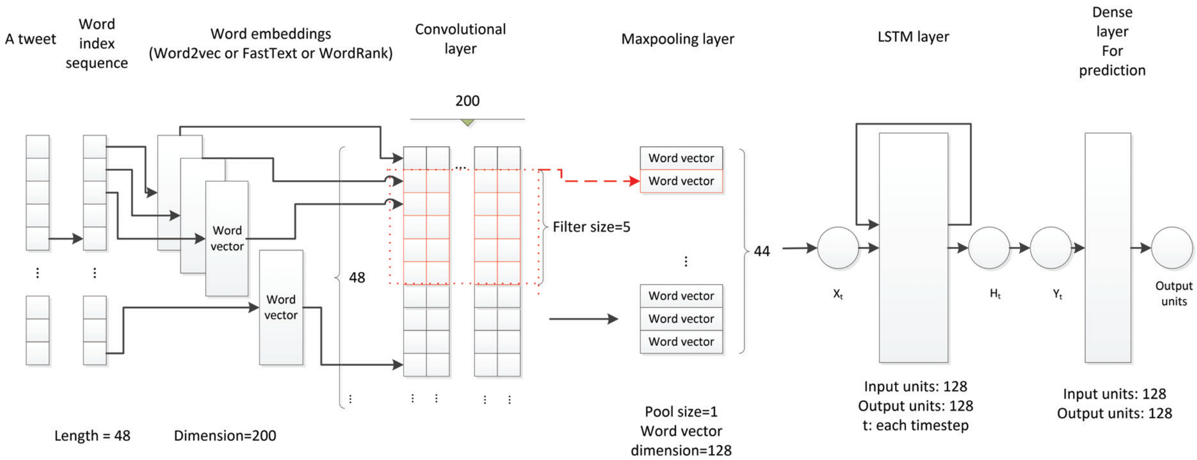

Figure 2: The architecture of convolutional LSTM neural network.

annotated. A set of 8,612 tweets were chosen as the training dataset and 3,719 tweets as the test dataset[4] . The information of the datasets is shown as Table 1.

Table 1: Training and test data sets

	Training Set			Test Set		
	# of Tweets	# of PETs	# of Non-PETs	# of Tweets	# of PETs	# of Non-PETs
Count	8,612	2,065	6,547	3,719	897	2,822

Results and Discussion

To understand how well our approach performs, four conventional classifiers were chosen to generate the baseline classification results, and they are logistic regression, decision tree, kNN and support vector machine. In our experiment, these classifiers were fed with 22 features engineered by researchers of our group. These features include POS tags, count of URLs, user counts, gramulator features, and screen name based features. For a detailed description of these features see (Jiang, Calix, and Gupta 2016). On the other hand, we fed the output of the convolutional network to our LSTM classifier. To investigate the performance of different word embedding models, we experimented with 3 different word embedding models discussed previously.

The result of our experiment is shown in Table 2. The table shows a comparison of the performance of the baseline classifiers with human engineered features versus the performance of the LSTM based deep neural network with word embedding features.

Results (Table 2) demonstrate that our approach to detect the PETs with LSTM based deep neural network with word embedding features outperforms, in all aspects, baseline classifiers that used human engineered features. In other

[4]The annotated data set named Medicine Corpus are available at: https://github.com/medeffects/tweet_corpora

Table 2: Performance comparison of classifying PETs between conventional classifiers with 22 engineered features and the convolutional LSTM with word index vectors

Input		classifier	Acc	Prec	Recall	F1	ROC
22 Engineered Features		**Logistic Regression**	0.660	0.341	0.452	0.389	0.574
		Decision Tree	0.620	0.305	0.463	0.368	0.551
		KNN	0.640	0.337	0.511	0.406	0.575
		SVM	0.660	0.338	0.441	0.383	0.571
Word Index Vectors	**word2vec**	**Conv. LSTM (normal)**	0.824	0.645	0.601	0.622	0.748
		Conv. LSTM (class weight)	0.814	0.600	0.746	0.665	0.778
	fastText	**Conv. LSTM (normal)**	**0.874**	**0.762**	0.695	**0.726**	0.826
		Conv. LSTM (class weight)	0.843	0.687	0.641	0.663	0.774
	WordRank	**Conv. LSTM (normal)**	0.828	0.645	0.643	0.644	0.765
		Conv. LSTM (class weight)	0.853	0.667	**0.796**	**0.726**	**0.832**

Table 3: Improvement of the best performance measures.

	Conventional classifiers with engineered features	Conv. LSTM with word embeddings	% Change
Acc	0.660	0.874	32
Prec	0.341	0.762	123
Recall	0.511	0.796	56
F1	0.406	0.726	79
ROC	0.575	0.832	45

words, unsupervised learned features seem to represent the semantics embedded in the tweet text better than engineered features. This can significantly reduce the efforts needed to engineer features.

Considering the best performances in each approach, one can see that accuracy has improved from 0.660 to 0.874, precision from 0.341 to 0.762, recall from 0.511 to 0.796, F1 from 0.406 to 0.726, and ROC from 0.575 to 0.832 (see Table 3). The biggest improved measure is the precision. This is of significant importance to this work, because a high precision implies a high true positive rate and low false positive rate–that is, more actual PETs (true positives) will be correctly identified in the predicted PETs (true positives + false positives). Achieving high precision has always been a goal in detecting PETs.

Another intriguing observation from the results in Table 2 is that in all performance measures, the most popular word embedding model, word2vec, does not perform the best. This may be a surprise, but it may confirm with the testing results by Rare Technology (Sethi 2017): fastText and WordRank have higher semantic accuracy.

Conclusion

We investigated an approach to combine word embedding techniques with convolutional and LSTM deep neural networks to detect PETs from Twitter data. We treated tweets as matrices of size 48x200, and used them with a CNN for salient feature detection. After CNN processing, better quality feature representations were fed in sequence to an LSTM based classifier. The results show that the proposed method outperforms the conventional classifiers which use human engineered features. The fastText and WordRank word vector space models have shown their advantages in providing a CNN with an excellent word vector space representation.

Compared with conventional classifiers, the combination of word embedding techniques and convolutional LSTM neural networks is not only a more accurate method to detect PETs, but can also accelerate the development process by not using human engineered features. For health surveillance, an efficient methodology such as the one proposed in this paper is crucial to deal with large scale social media data.

Acknowledgment

Authors wish to thank Yongbing Tang, Dustin Franz, Ravish Gupta for collecting the Twitter data, Alexandra Vest, Cecelia Lai, Bridget Swindell, and Mary Stroud for annotating the tweets, and Liyuan Huang and Tae-Hoon Kim for guiding with hardware/GPU support. This work was supported in part by the National Institutes of Health Grant 1R15LM011999-01.

References

Bilge, U.; Bozkurt, S.; Yolcular, B. O.; and Ozel, D. 2012. Can social web helpoff detect influenza related illnesses in turkey? *Stud Health Technol Inform* 174:100–104.

Bojanowski, P.; Grave, E.; Joulin, A.; and Mikolov, T. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Broniatowski, D. A.; Paul, M. J.; and Dredze, M. 2013. National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. *PloS one* 8(12):e83672.

Chary, M.; Genes, N.; McKenzie, A.; and Manini, A. F. 2013. Leveraging social networks for toxicovigilance. *Journal of Medical Toxicology* 9(2):184–191.

Chew, C., and Eysenbach, G. 2010. Pandemics in the age of twitter: content analysis of tweets during the 2009 h1n1 outbreak. *PloS one* 5(11):e14118.

Chunara, R.; Andrews, J. R.; and Brownstein, J. S. 2012. Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak. *The American journal of tropical medicine and hygiene* 86(1):39–45.

Collier, N.; Son, N. T.; and Nguyen, N. M. 2011. Omg u got flu? analysis of shared health messages for bio-surveillance. *Journal of biomedical semantics* 2(5):S9.

Coloma, P. M.; Becker, B.; Sturkenboom, M. C.; Van Mulligen, E. M.; and Kors, J. A. 2015. Evaluating social media networks in medicines safety surveillance: two case studies. *Drug safety* 38(10):921–930.

Dredze, M. 2012. How social media will change public health. *IEEE Intelligent Systems* 27(4):81–84.

Elman, J. L. 1990. Finding structure in time. *Cognitive science* 14(2):179–211.

Freifeld, C. C.; Brownstein, J. S.; Menone, C. M.; Bao, W.; Filice, R.; Kass-Hout, T.; and Dasgupta, N. 2014. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug safety* 37(5):343–350.

Fung, I. C.-H.; Fu, K.-W.; Ying, Y.; Schaible, B.; Hao, Y.; Chan, C.-H.; and Tse, Z. T.-H. 2013. Chinese social media reaction to the mers-cov and avian influenza a (h7n9) outbreaks. *Infectious diseases of poverty* 2(1):31.

Gers, F. A.; Schmidhuber, J.; and Cummins, F. 1999. Learning to forget: Continual prediction with lstm.

Gesualdo, F.; Stilo, G.; Gonfiantini, M. V.; Pandolfi, E.; Velardi, P.; Tozzi, A. E.; et al. 2013. Influenza-like illness surveillance on twitter through automated learning of naïve language. *PLoS One* 8(12):e82489.

Hanson, C. L.; Burton, S. H.; Giraud-Carrier, C.; West, J. H.; Barnes, M. D.; and Hansen, B. 2013. Tweaking and tweeting: exploring twitter for nonmedical use of a psychostimulant drug (adderall) among college students. *Journal of medical Internet research* 15(4).

Harris, J. K.; Mansour, R.; Choucair, B.; Olson, J.; Nissen, C.; Bhatt, J.; for Disease Control, C.; Prevention; et al. 2014. Health department use of social media to identify foodborne illness-chicago, illinois, 2013-2014. *MMWR Morb Mortal Wkly Rep* 63(32):681–685.

Heaivilin, N.; Gerbert, B.; Page, J.; and Gibbs, J. 2011. Public health surveillance of dental pain via twitter. *Journal of dental research* 90(9):1047–1051.

Jashinsky, J.; Burton, S. H.; Hanson, C. L.; West, J.; Giraud-Carrier, C.; Barnes, M. D.; and Argyle, T. 2014. Tracking suicide risk factors through twitter in the us. *Crisis*.

Jiang, K., and Zheng, Y. 2013. Mining twitter data for potential drug effects. In *International Conference on Advanced Data Mining and Applications*, 434–443. Springer.

Jiang, K.; Tang, Y.; Cook, G. E.; and Madden, M. M. 2017. Discovering potential effects of dietary supplements from twitter data. In *Proceedings of the 2017 International Conference on Digital Health*, 119–126. ACM.

Jiang, K.; Calix, R.; and Gupta, M. 2016. Construction of a personal experience tweet corpus for health surveillance. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, 128–135.

Lai, S.; Xu, L.; Liu, K.; and Zhao, J. 2015. Recurrent convolutional neural networks for text classification. In *AAAI*, volume 333, 2267–2273.

Lev, G.; Klein, B.; and Wolf, L. 2015. In defense of word embedding for generic text representation. In *International Conference on Applications of Natural Language to Information Systems*, 35–50. Springer.

McManus, K.; Mallory, E. K.; Goldfeder, R. L.; Haynes, W. A.; and Tatum, J. D. 2015. Mining twitter data to improve detection of schizophrenia. *AMIA Summits on Translational Science Proceedings* 2015:122.

Mikolov, T.; Karafiát, M.; Burget, L.; Cernockỳ, J.; and Khudanpur, S. 2010. Recurrent neural network based language model. In *Interspeech*, volume 2, 3.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Nagar, R.; Yuan, Q.; Freifeld, C. C.; Santillana, M.; Nojima, A.; Chunara, R.; and Brownstein, J. S. 2014. A case study of the new york city 2012-2013 influenza season with daily geocoded twitter data from temporal and spatiotemporal perspectives. *Journal of medical Internet research* 16(10).

Nagel, A. C.; Tsou, M.-H.; Spitzberg, B. H.; An, L.; Gawron, J. M.; Gupta, D. K.; Yang, J.-A.; Han, S.; Peddecord, K. M.; Lindsay, S.; et al. 2013. The complex relationship of realspace events and messages in cyberspace: case study of influenza and pertussis using tweets. *Journal of medical Internet research* 15(10).

Nascimento, T. D.; DosSantos, M. F.; Danciu, T.; DeBoer, M.; van Holsbeeck, H.; Lucas, S. R.; Aiello, C.; Khatib, L.; Bender, M. A.; et al. 2014. Real-time sharing and expression of migraine headache suffering on twitter: a cross-sectional infodemiology study. *Journal of medical Internet research* 16(4).

Odlum, M., and Yoon, S. 2015. What can we learn about the ebola outbreak from tweets? *American journal of infection control* 43(6):563–571.

Sarker, A.; Ginn, R.; Nikfarjam, A.; OConnor, K.; Smith, K.; Jayaraman, S.; Upadhaya, T.; and Gonzalez, G. 2015. Utilizing social media data for pharmacovigilance: A review. *Journal of biomedical informatics* 54:202–212.

Sethi, P. 2017. Wordrank embedding: crowned is most similar to king, not word2vecs canute. https://rare-technologies.com/wordrank-embedding-crowned-is-most-similar-to-king-not-word2vecs-canute/.

Severyn, A., and Moschitti, A. 2015. Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 959–962. ACM.

Signorini, A.; Segre, A. M.; and Polgreen, P. M. 2011. The use of twitter to track levels of disease activity and public concern in the us during the influenza a h1n1 pandemic. *PloS one* 6(5):e19467.

Sofean, M., and Smith, M. 2012. Sentiment analysis on smoking in social networks. *Studies in health technology and informatics* 192:1118–1118.

Public health surveillance. http://www.who.int/topics/public_health_surveillance/en/.

Wu, Y.-C.; Yin, F.; and Liu, C.-L. 2017. Improving handwritten chinese text recognition using neural network language models and convolutional neural network shape models. *Pattern Recognition* 65:251–264.

Yepes, A. J.; MacKinlay, A.; and Han, B. 2015. Investigating public health surveillance using twitter. *ACL-IJCNLP* 2015:164.

Yom-Tov, E.; Borsa, D.; Cox, I. J.; and McKendry, R. A. 2014. Detecting disease outbreaks in mass gatherings using internet data. *Journal of medical Internet research* 16(6).

Zhang, X.; Zhao, J.; and LeCun, Y. 2015. Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, 649–657.