RESEARCH

Neural attention with character embeddings for hay fever detection from twitter

Jiahua Du*, Sandra Michalska, Sudha Subramani, Hua Wang and Yanchun Zhang

Abstract

The paper aims to leverage the highly unstructured user-generated content in the context of pollen allergy surveillance using neural networks with character embeddings and the attention mechanism. Currently, there is no accurate representation of hay fever prevalence, particularly in real-time scenarios. Social media serves as an alternative to extract knowledge about the condition, which is valuable for allergy sufferers, general practitioners, and policy makers. Despite tremendous potential offered, conventional natural language processing methods prove limited when exposed to the challenging nature of user-generated content. As a result, the detection of *actual* hay fever instances among the number of false positives, as well as the correct identification of non-technical expressions as pollen allergy symptoms poses a major problem. We propose a deep architecture enhanced with character embeddings and neural attention to improve the performance of hay fever-related content classification from Twitter data. Improvement in prediction is achieved due to the character-level semantics introduced, which effectively addresses the out-of-vocabulary problem in our dataset where the rate is approximately 9%. Overall, the study is a step forward towards improved real-time pollen allergy surveillance from social media with state-of-art technology.

Keywords: Pollen allergy, Hay fever, Twitter, Deep learning

Introduction

Nearly one in five Australians suffered from hay fever in between 2014 and 2015 [1]. According to the World Health Organization [35], pollen allergy will only increase in prevalence and severity over the next decade, which leads to a global concern. Unsurprisingly, the accurate estimates of hay fever remain the top priority for Australian Institute of Health and Welfare. The traditional data sources for the condition scale evaluation include official statistics, general practitioner records, hospital admissions [34], antihistamine sales, etc. Still, the substantial time lag in the results reporting as well as the insufficient data granularity do not allow to obtain the accurate representation of pollen allergy prevalence and severity in real-time. Recently, social media data mining for public health surveillance has been growing in popularity in the research communities to account for the limitations of the existing methods. The critical challenge though lies in extracting the relevant information from highly unstructured data.

Social media platforms in particular abound in misspellings, abbreviations, and informal expressions. Neural network approaches have been applied lately in place of conventional machine learning techniques [17, 22, 25, 36] to account for the syntactic and semantic relationships between the words. Attention mechanism further enables to discriminate the words/phrases that contribute mostly towards the respective class assignment. To illustrate the problem, below presents the examples of relevant versus non-relevant posts, despite similar wording causing confusion in classification.

1. Hay fever-**related** content (Symptoms)
   My **eyes have been watering** and **I've been sneezing** heaps today... anyone else in Melbourne noticing their hay fever kicking in for the first time this spring?
2. Hay fever-**non-related** content (Advertisement)
   New one-off treatment for hay fever sufferers. It reduces symptoms by freezing the nerves that make you **sneeze**.

*Correspondence: jiahua.du@live.vu.edu.au
Institute of Sustainable Industries & Liveable Cities, Victoria University, Melbourne, VIC, Australia

Du *et al. Health Inf Sci Syst (2019) 7:21*

Page 2 of 7

Traditional rule-based classifiers would consider the post containing term '*sneeze*' relevant to the case study, thus overestimating the actual hay fever prevalence. On the other hand, numerous posts include pollen allergy-related terms expressed in the non-medical jargon, which proves impossible to detect using the pre-defined rules. Furthermore, the numerous Out-Of-Vocabulary (OOV) words prevalent on social media platforms significantly reduce the classification performance due to their absence in training dataset.

The primary objective of our study is to implement and validate the most recent neural networks approach with attention mechanism and character embeddings to the challenging problem of hay fever detection from Twitter. The posts are automatically classified into the 4 pre-defined classes (see Sect. 3.2) based on their relevance. Neural attention allows to increase the weight of the constituent parts of the posts that play a major role in final class prediction. The character embeddings on the other hand effectively tackle the OOV problem by utilizing the similar words syntax. We demonstrate an improvement in accuracy and macro-F1 of our proposed model as well as illustrate that our approach can both (i) capture the hay fever-related information in highly informal tweets (e.g. Symptoms, Treatments), and (ii) identify the non-relevant content (e.g. Advertisements, Warnings). The study provides the proof-of-concept of state-of-the-art techniques application in the context of pollen allergy towards improved public health surveillance from social media.

Related work
Health surveillance from social media
Individuals often prefer to share health-related experiences with peers, rather than during clinical studies, or even physicians [10]. In addition, the knowledge based solely on health practitioners' reports and patients' surveys tend to be generic and often limited in scope [8]. Furthermore, Cvetkovski et al. [9] reported that hay fever tend to be self-managed and availability of over-the-counter medications leads to bypassing the health care professionals, putting additional pressure on complementary data sources about the condition surveillance.
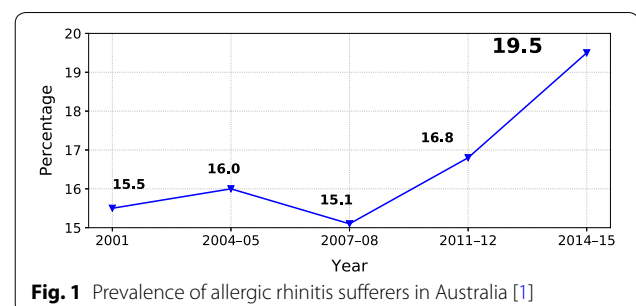
Given the limitations, social media has opened an enormous opportunity for public health surveillance from directly affected users. In particular, Twitter has recorded approximately three million active accounts since January 2018 in Australia [7]. Due to its short format, Twitter also encourages the high frequency of updates [18]. This in turn generates an abundance of data, commonly concerning the health-related matters [5]. As a result, Twitter has drawn attention from public health communities to answer numerous health-related questions [2].

In the case of pollen allergy surveillance, De Quincey et al. [11, 12] demonstrated that Twitter enables researchers to access information regarding the specific pollen allergy symptoms, as well as the medications usage and effectiveness. The comparison with UK Pollen Hotzones further proved that geolocated Twitter data is a good proxy for the condition prevalence estimation due to the similar distribution [12]. In the other study, Gesualdo et al. [14] observed the high correlation between pollen counts and tweets reporting hay fever incidents in the study conducted in the US. The results obtained serve as a proof of concept of the potential role of social media in signalling allergic symptoms and drug consumption trends.

Hay fever prevalence in Australia
Three million Australian adults struggle through spring and summer with symptoms such as watery eyes, running nose, itchy throat, sneezing or irritability. Pollen allergy is also considered the most common chronic respiratory disease in Australia [1], posing a significant health and economic burden [9]. The quality of life of allergy sufferers is substantially reduced, affecting physical, psychological, and social functioning [35]. According to the National Health Survey conducted by the Australian Bureau of Statistics, which is shown in Fig. 1, the prevalence of allergic rhinitis among Australians has been measured over the past 15 years and indicated growth over time.

The exact estimations of hay fever prevalence proves a challenging task due to the limited resources, i.e. time- and cost-consuming official statistics, marketing surveys, pharmaceuticals data, etc. The usual peak in hay fever occurrences is observed around spring and summer period. However, climate changes observed are lengthening the pollen seasons as well as introducing an increased intensity of allergens, and unexpected new pollens in certain areas [35]. Additionally, the increasing air pollution, especially around urban areas further affects the respiratory health of the population. This in turn adds an uncertainty to the accurate hay fever prevalence estimation.

Fig. 1 Prevalence of allergic rhinitis sufferers in Australia [1]

Du *et al. Health Inf Sci Syst* (2019) 7:21

Page 3 of 7

The real-time monitoring proves invaluable for allergy sufferers, health practitioners, and policy makers.

Deep learning in text classification

Previous studies on allergies surveillance from social media conducted in UK and US utilized either traditional machine learning classifiers, including Naive Bayes [6, 24] and lexicon-based approaches [14, 11, 12]. Despite the wealth of knowledge that social media offers, the natural language used still constitutes a major challenge in tweets analysis, and forms an obstacle in relevant information extraction [29]. For instance, the highly informal and continuously evolving vocabulary such as '*dribbling nose*' and '*hay fever sob*' prove difficult to classify as the potential symptoms and map to their medical equivalents, i.e. '*runny nose*', '*watery eyes*'. Lack of advanced Natural Language Processing (NLP) techniques addressing the above mentioned issues leads to the limited applicability of the approaches in the case of the emerging symptoms/treatments, not identified a priori. Despite the existing shortcomings, no previous study applied deep learning to user-generated content classification in the context of hay fever. Furthermore, the performance improvement using neural attention is yet to be discovered in literature.

Deep learning has already proven successful in text classification tasks, outperforming the conventional machine learning techniques [16, 21, 26, 27], effectively capturing both syntactic (e.g. *allergy, allergic, allergen,* etc.) and semantic (e.g. *hay fever, pollen allergy, allergic rhinitis*) word dependencies. Also, deep learning alleviates the need for laborious and time-consuming manual feature engineering. The most distinctive features are extracted automatically from the raw input during the model training. The successful application of deep learning has been reported in numerous NLP tasks, including topic categorization [19], machine translation [31], sentence modelling [20], and Part-Of-Speech tagging [4].

Among many neural architectures, Recurrent Neural Networks (RNNs), in particular Long Short-Term Memory networks (LSTMs) [15] are widely implemented to model text sequences due to their capability in modeling long-range dependencies and historical information storage over time [32]. Attention mechanism can further boost the performance of RNNs by focusing on the time-steps that are most critical to the task [15]. In regular RNNs (without attention), the prediction is made using RNNs at the final time-step. With attention, RNNs save the output at every time-step, and the mechanism then selects and combines the most important outputs based on their relevance to the task [13]. The improved performance of RNNs with attention versus RNNs without attention was obtained in the case study of information extraction from cancer pathology reports [13].

Methodology

Data extraction

The data was sourced from the micro-blogging platform Twitter commonly used for public health surveillance purposes due to the high frequency of updates, convenience of geo-location, and the wide API availability. The geo-coordinates were set to the latitude and longitude of Alice Springs (centre of Australia), and the radius of 2000 miles. The tweets were extracted weekly using the R programming language and TwitteR package. The search criteria were 'hayfever' and 'hay fever' in order to ensure the high precision over recall of the returning tweets. The collection period was from June 1, 2018 to December 31, 2018, and covered the high pollen season in Australia. The data was obtained for 191 out of 214 days in total (89%). The tweets from the remaining 23 days were not captured due to technical issues. The mention of Twitter users are removed to ensure compliance to privacy and ethical considerations.

Data annotation

The entire dataset containing 4148 posts was annotated by two researchers from health informatics. The annotation process followed the schema presented in Table 1, with tweet examples representative of each class. The respective categories were based on the relevancy to the aim of the study (1-most relevant, 4-least relevant). The two over-arching groups were distinguished, i.e. Informative and Non-Informative. The Cohen's kappa statistic [3] was calculated to measure the inter-rater reliability. The score of $\kappa = 0.82$ was obtained, which is considered significant based on the study by Viera et al. [33]. Any potential disagreements were either resolved by consensus, otherwise the 'Unrelated/Ambiguous' class was assigned.

Data pre-processing

First, retweets and duplicate tweets are removed to prevent the dataset from redundant information. The remaining tweets are then tokenized into word sequences via a twitter-aware tokenizer. For each tokenized tweet: (i) all words are lowercased with the constituent URLs, punctuation, and stopwords removed; (ii) for each word, any characters repeating over three times in a word will be normalized by removing the remaining repetitions. As a result, the pre-processed dataset consists of 3862 tweets, covering 719 samples for Class 1, 1823 samples for Class 2, 938 samples for Class 3, and 382 samples for Class 4.

Models for real-time hay fever detection

The real-time hay fever detection on tweets is regarded as a multi-class classification task. Given a set of tweets *D*, each

Table 1 Annotation schema with the examples of tweets

Class	Description	Example
Informative		
1	Detailed personal reporting (symptoms, treatments, etc.)	My eyes have been watering and I've been sneezing heaps today...anyone else in Melbourne noticing their hay fever kicking in for the first time this spring?
2	Generic personal reporting	I wanted a Sunday morning lie-in, but hayfever is telling me different
Non-informative		
3	Warnings/news/marketing	Struggling with athsma or hayfever? Find out how a #saltlamp can help
4	Ambiguous/un-related	If I had hayfever I would simply buy some hay

tweet $d \in D$ is associated with one of the four predefined classes $y = \{1, 2, 3, 4\}$. The goal of the task is to learn automatically representative features that are able to correctly predict the corresponding class of a tweet. The architecture is presented as follows.

The architecture starts by encoding word-level semantic information of tweets. Let a tweet $d \in D$ be a sequence of n words $d = \{x_1, x_2, \ldots, x_n\}$. Each word $x \in V$ is associated with a d-dimensional embedding $\mathbf{e}_x$ to encode word-level semantics, which is stored in a word lookup table $\mathbf{E} \in \mathbb{R}^{|V| \times d}$, where V is the vocabulary of D.

In addition, character-level semantics is utilized to alleviate the OOV problem. Following [37], constituent characters of a word $x \in V$ are embedded into vectors, which are used to learn a l-dimensional character embedding $\mathbf{e}'_x$ encoding sub-word information of the word via convolutional neural networks [23]. Each word x thus can be represented by concatenating both the corresponding word embedding $\mathbf{e}_x$ and character embedding $\mathbf{e}'_x$. The representation of a tweet d is obtained by stacking its words:

$$\mathbf{X} = [\mathbf{e}_{x_1} \oplus \mathbf{e}'_{x_2}, \mathbf{e}_{x_1} \oplus \mathbf{e}'_{x_2}, \ldots, \mathbf{e}_{x_n} \oplus \mathbf{e}'_{x_n}], \tag{1}$$

where $\oplus$ is the concatenation operator, the matrix $\mathbf{X} \in \mathbb{R}^{n \times (d+l)}$ represents a tweet, serving as one input sample for model fitting.

Given an input sample $\mathbf{X}$, a bidirectional LSTM is employed to obtain word annotations by summarizing contextual information of each word in a tweet.

$$[\overrightarrow{\mathbf{h}}_1, \overrightarrow{\mathbf{h}}_2, \ldots, \overrightarrow{\mathbf{h}}_n] = \overrightarrow{\mathrm{LSTM}}(\mathbf{X}), \tag{2}$$

$$[\overleftarrow{\mathbf{h}}_1, \overleftarrow{\mathbf{h}}_2, \ldots, \overleftarrow{\mathbf{h}}_n] = \overleftarrow{\mathrm{LSTM}}(\mathbf{X}). \tag{3}$$

The final annotation of a word $\mathbf{h} = \overrightarrow{\mathbf{h}} \oplus \overleftarrow{\mathbf{h}}$ results from concatenating the forward hidden state $\overrightarrow{\mathbf{h}} \in \mathbb{R}^m$ and backward hidden state $\overleftarrow{\mathbf{h}} \in \mathbb{R}^m$. The sentence embedding $\mathbf{s}$ representing the tweet is then constructed using the annotations via attention mechanisms:

$$\mathbf{z}_i = \tanh(\mathbf{W}_a^\top \mathbf{h}_i + \mathbf{b}_a), \tag{4}$$

$$\alpha_i = \frac{\exp(\mathbf{z}_i^\top \hat{\mathbf{h}})}{\sum_{j=1}^n \exp(\mathbf{z}_j^\top \hat{\mathbf{h}})}, \tag{5}$$

$$\mathbf{s} = \sum_{i=1}^n \alpha_i \mathbf{h}_i, \tag{6}$$

where $\mathbf{W}_a$, $\mathbf{b}_a$, and $\hat{\mathbf{h}}$ are learned parameters. The sentence embedding $\mathbf{s} \in \mathbb{R}^m$ is forwarded into an output layer to predict the probability distribution $\mathbf{p}$ over the predefined classes.

$$\mathbf{p} = \mathrm{softmax}(\mathbf{W}_o^\top \mathbf{s} + b_o). \tag{7}$$

The model is trained to minimize the cross entropy between true labels and the predicted labels across tweets in the dataset D.

$$\mathcal{L} = -\sum_{d \in D} \log \mathbf{p}_d(y_d), \tag{8}$$

where y_d is the true label and $\mathbf{p}_d(y_d)$ is the probability of the true label. For simplicity, the attention-based bidirectional LSTM model is henceforth called BILSTM + ATT, whereas our architecture BILSTM + ATT + CHAR.

Model training and evaluation

The neural models are evaluated using stratified 10-fold cross validation. As Table 2 shows, the dataset has on average approximately 9% missing words across testing folds, whereas the rate of missing characters is far lower.

The following hyper-parameters are used. The dimensionality of word and character embeddings $d = l = 50$. The word lookup table $\mathbf{E}$ is initialized with the GloVe embeddings pre-trained [28] on two billion tweets. The number and region size of kernels used to extract character embeddings are set to 100 and 5, respectively. The number of LSTM units is set to $m = 50$. For each fold, 10% of the training data is randomly selected for

Du *et al. Health Inf Sci Syst (2019) 7:21*

Page 5 of 7

Table 2 OOV rate of words and characters across testing folds

Fold	1	2	3	4	5	6	7	8	9	10	Average
OOV words	8.64	8.52	9.26	8.19	9.42	9.08	9.79	7.57	8.88	9.89	8.92
OOV characters	0.07	0.07	0.03	0.06	0.03	0.05	0.01	0.04	0.03	0.06	0.04

Table 3 The performance of model variants

Model	Accuracy	Macro-F1
BILSTM + ATT	77.72	72.80
BILSTM + ATT + CHAR	79.51	75.67

validation, with mini-batch size of 64 and the Adam optimizer updating neural weights. Early stopping is employed to prevent over-fitting when the validation loss stops improving for 10 epochs. Each model is trained 5 times to report average results for model robustness under random parameter initialization.

Result analysis

Quantitative evaluation

Both accuracy and macro-F1 are used to evaluate model performance on the dataset with imbalanced labels. Table 3 demonstrates that incorporating character embeddings that encode sub-word information outperforms the original attention-based model by about 2% in accuracy. The improvement in macro-F1 is even more with almost 3%, suggesting that sub-word information is useful in hay fever prediction for classes with fewer training samples.

Qualitative analysis

Attention mechanism

has the advantage of generating higher weights to the words that contribute mostly towards the relevant class assignment. Table 4 produces examples of posts with the attention maps overlaid on the text. Higher color intensity indicates greater attention score, and thus more impact on the final class prediction. For instance, in the posts from Class 1 (Symptoms/Treatments), the higher weight was assigned to the words 'crying' (= watery eyes), and 'meds' (= medications), despite no obvious symptom/treatment indication. In the case of Class 2 (Generic reporting), hay fever-related terms (e.g. 'season', 'pollen') were attended more in model learning. Also, the term 'windy' obtained a higher score, which can be considered intuitive as pollen spreads more intensively during strong winds. As for Class 3 (Warnings/News/Marketing), the more formal and non allergy-related words such as 'surge' and 'news' mostly contributed to the News/Warnings prediction. Finally, in the posts from Class 4 (Ambiguous/Un-related) the irrelevant terms from hay fever detection point of view were highlighted, i.e. the words 'parliament' or 'bottle' were un-seen in the hay fever-related content, thus became decisive of Ambiguous/Unrelated class assignment.

Table 4 The examples of post with the attention maps

Class	Tweets with Attention Map
1	(1) i'm not crying it's my hay fever acting up
	(2) took some hayfever meds last night to help me sleep and holy shit do people with seasonal allergies get to feel like
2	(1) hayfever season is here pollen on my car today and i feel like absolute crap
	(2) it's windy rainy dusty and hot i am absolutely about to get fucked over by my hayfever
3	(1) surge in hay fever is rooted in our modern lifestyles
	(2) great news for anyone who suffers from asthma hay fever or allergies
4	(1) i'm going to run for state parliament on platform of promising to cut down every plane tree in melbourne hay fever sucks
	(2) who genius who decided to bottle air from country with one of highest hay fever and tree in asthma rates in the world

Color intensity indicates the weight attributed to each word towards the respective class assignment

Table 5 The examples of posts with OOV words and their respective predictions probabilities for BILSTM + ATT (A) and BILSTM + ATT + CHAR (B)

Class	Prob. (A)	Prob. (B)	Post
1	0.60	0.99	**um** i'm still sick also hay fever kicking **n** eyes **r** burning **n** red day **v** self
1	0.97	0.98	**aaahhh** hayfever season allow play song people ...sneezes minutes straight
1	0.20	0.99	seriously one bizarre moments life told buy usual effective hayfever **med**
2	0.03	0.50	**probs** hay fever
2	0.94	1.00	canny sun **coz** hay fever
2	0.80	1.00	im thinking immune **aka** hayfever due pollen air last couple days
4	0.26	0.67	last one like **ads** hayfever sufferers going mars avoid allergens

Character embeddings

encode character-level features of the words that share similar prefix/suffix [30], and obtain closer representation among words within the same class based on their characters [37]. The implementation of character embeddings is mainly driven by the occurrence of OOV problem, which is particularly visible in the case of social media data due to the number of grammatical errors (e.g. 'athsma'), abbreviated forms (e.g. 'meds'), jargonic expressions (e.g. 'sniffling'), etc.

In Table 5, the examples of posts with OOV have been illustrated, including the respective class probabilities obtained for both models. The probabilities for the model with character embeddings were higher, demonstrating greater confidence towards the actual class prediction. This can be attributed to the OOV recognition and mapping, where similar syntactical structure was utilized, i.e. **med** = **med**ication, **ads** = **ad**vertisement**s**, **probs** = **prob**ably, etc. Such an approach proves particularly beneficial in the context of health surveillance, where complex medical terminology (e.g. Allergic Rhinitis), treatment names (e.g. antihistamines), and medication brands (e.g. Loratadine) are highly prone to frequent mispronunciation on social media platforms.

Conclusions

Pollen allergy is a major health and economic burden, impacting the lives of approximately 20% of Australian population, in particular working-aged adults. The real-time monitoring of the condition prevalence and severity is currently unavailable. Twitter data has been proven to be a valuable source of information on emerging symptoms as well as treatments usage from directly affected individuals. The challenging nature of user-generated content (i.e. misspellings, abbreviations, jargon) poses significant limitation in actionable knowledge extraction. Deep learning is currently the state-of-the-art in NLP tasks. We employ the neural networks model with the attention mechanism to account for the fact that each word contributes differently towards final class assignment. We further improve the classification accuracy of the model with character embeddings implementation, which effectively addresses the OOV problem. We also discuss the inner workings of the attention mechanism as well as character embeddings on the examples of posts to facilitate finding interpretability. Taken together, the study demonstrates and validates the practical application of state-of-art deep learning in the context of pollen allergy surveillance from social media.

References

1. Australian Institute of Health and Welfare (AIHW). Allergic rhinitis ('hay fever'). https://www.aihw.gov.au/reports/chronic-respiratory-conditions/allergic-rhinitis-hay-fever/contents/allergic-rhinitis-by-the-numbers (2016). Accessed 30 Jan 2019.
2. Byrd K, Mansurov A, Baysal O. Mining twitter data for influenza detection and surveillance. In: Proceedings of the international workshop on software engineering in healthcare systems. New York: ACM; 2016. p. 43–9.
3. Carletta J. Assessing agreement on classification tasks: the kappa statistic. Comput linguist. 1996;22(2):249–54.
4. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. J Mach Learn Res. 2011;12(Aug):2493–537.
5. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in twitter. In: Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality, 2014, p. 51–60.
6. Cowie S, Arthur R, Williams H. @ choo: tracking pollen and hayfever in the UK using social media. Sensors. 2018;18(12):4434.
7. Cowling D. Social media statistics Australia—January 2018. 2018. https://www.socialmedianews.com.au/social-media-statistics-australia-january-2018. Accessed 29 June 2019.
8. Culotta A. Estimating county health statistics with twitter. In: Proceedings of the 32nd annual ACM conference on Human factors in computing systems. New York: ACM; 2014. p. 1335–44.
9. Cvetkovski B, Kritikos V, Yan K, Bosnic-Anticevich S. Tell me about your hay fever: a qualitative investigation of allergic rhinitis management from the perspective of the patient. NPJ Prim Care Respir Med. 2018;28(1):3.
10. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. Am Psychol. 2000;55(2):205.
11. de Quincey E. Potential of social media to determine hay fever seasons and drug efficacy. Planet Risk. 2014;2(4):293–7.

Du *et al. Health Inf Sci Syst (2019) 7:21*

Page 7 of 7

12. de Quincey E, Kyriacou T, Pantin T. # hayfever; a longitudinal study into hay fever related tweets in the UK. In: Proceedings of the 6th international conference on digital health conference. New York: ACM; 2016, p. 85–9.

13. Gao S, Young MT, Qiu JX, Yoon HJ, Christian JB, Fearn PA, Tourassi GD, Ramanthan A. Hierarchical attention networks for information extraction from cancer pathology reports. J Am Med Inf Assoc. 2017;25(3):321–30.

14. Gesualdo F, Stilo G, D'Ambrosio A, Carloni E, Pandolfi E, Velardi P, Fiocchi A, Tozzi AE. Can twitter be a source of information on allergy? Correlation of pollen counts with tweets reporting symptoms of allergic rhinoconjunctivitis and names of antihistamine drugs. PloS ONE. 2015;10(7):e0133706.

15. Graves A. Generating sequences with recurrent neural networks. 2013. arXiv:1308.0850.

16. Hu H, Li J, Wang H, Daggard G. Combined gene selection methods for microarray data analysis. In: International conference on knowledge-based and intelligent information and engineering systems. Berlin: Springer; 2006. p. 976–83.

17. Huang J, Peng M, Wang H, Cao J, Gao W, Zhang X. A probabilistic method for emerging topic tracking in microblog stream. World Wide Web. 2017;20(2):325–50.

18. Java A, Song X, Finin T, Tseng B. Why we twitter: understanding micro-blogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. New York: ACM; 2007. p. 56–65.

19. Johnson R, Zhang T. Effective use of word order for text categorization with convolutional neural networks. In: Human language technologies: the 2015 annual conference of the North American chapter of the ACL, Denver, CO, 2014, p. 103–12

20. Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, 2014, p. 655-65.

21. Khalil F, Li J, Wang H. An integrated model for next page access prediction. IJ Knowl Web Intell. 2009;1(1/2):48–80.

22. Khalil F, Wang H, Li J. Integrating markov model with clustering for predicting web page accesses. In: Proceeding of the 13th Australasian world wide web conference (AusWeb07). AusWeb; 2007, p. 63–74.

23. Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Doha, Qatar: Association for Computational Linguistics; 2014. p. 1746–51.

24. Lee K, Agrawal A, Choudhary A. Mining social media streams to improve public health allergy surveillance. In: 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE; 2015. p. 815–22.

25. Li H, Wang Y, Wang H, Zhou B. Multi-window based ensemble learning for classification of imbalanced streaming data. World Wide Web. 2017;20(6):1507–25.

26. Ma J, Sun L, Wang H, Zhang Y, Aickelin U. Supervised anomaly detection in uncertain pseudoperiodic data streams. ACM Trans Internet Technol (TOIT). 2016;16(1):4.

27. Peng M, Zeng G, Sun Z, Huang J, Wang H, Tian G. Personalized app recommendation based on app permissions. World Wide Web. 2018;21(1):89–104.

28. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, p. 1532–43.

29. Salloum SA, Al-Emran M, Monem AA, Shaalan K. A survey of text mining in social media: facebook and twitter perspectives. Adv Sci Technol Eng Syst J. 2017;2(1):127–33.

30. Santos CD, Zadrozny B. Learning character-level representations for part-of-speech tagging. In: Proceedings of the 31st international conference on machine learning (ICML-14), 2014, p. 1818–26.

31. Serban IV, Sordoni A, Bengio Y, Courville AC, Pineau J. Building end-to-end dialogue systems using generative hierarchical neural network models. 2016. arXiv:1507.04808

32. Subramani S, Michalska S, Wang H, Du J, Zhang Y, Shakeel H. Deep learning for multi-class identification from domestic violence online posts. IEEE Access. 2019;7:46210–24.

33. Viera AJ, Garrett JM, et al. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360–3.

34. Wang KN, Bell JS, Chen EYH, Gilmartin-Thomas JFM, Ilomäki J. Medications and prescribing patterns as factors associated with hospitalizations from long-term care facilities: a systematic review. Drugs Aging. 2018;35(5):423–57. https://doi.org/10.1007/s40266-018-0537-3.

35. World Allergy Organization (WAO). World allergy week 2016. 2016. https://www.worldallergy.org/UserFiles/file/WorldAllergyWeek2016FactSheet.pdf. Accessed 30 Jan 2019.

36. Zhang J, Tao X, Wang H. Outlier detection from large distributed databases. World Wide Web. 2014;17(4):539–68.

37. Zhang X, LeCun Y. Text understanding from scratch. 2015. arXiv:1502.01710.