on the information contribution related to the class variable, i.e. the amount of additional information about Z provided by X:

$$IG(Z;X) = H(Z) - H(Z|X) \tag{3}$$

After obtaining the information gain of each feature, a minimum limit was set to filter out the features whose information gain is below the limit.

2.5 Classification

After feature reduction, each news post was transformed into a number of features that are represented as numerals. This lays the basis for the application of machine learning methods [16-17]. Support vector machine (SVM) is one of the most popular machine learning algorithms for classification and regression problems. The core idea of the SVM is to separate different classes with a hyperplane. This algorithm has been proved to have excellent effect on linear data classification in many applications [18]. If the original data are only nonlinearly separable, the SVM can be coupled with kernel functions to map the original data the feature space to a higher-dimensional space, such as to separate the data linearly. This subsection briefly describes the SVM algorithm adopted for our binary classification problem.

For linearly separable data, the decision function can be written as:

$$\mathrm{f(x)} = \mathrm{wTx} + \varepsilon = 0 \tag{4}$$

where, w is the weight vector; ε is the bias; x is the dataset. The hyperplane described by formula (4) divides one space into two parts: a positive part for samples in the positive class (+) and a negative part for samples in the negative class (-). Since the problem is to determine the values of w and ε, so that the hyperplane can be as far as possible from all the samples. More specifically, the SVM algorithm sets up hyperplanes, HP_1 and HP_2, as follows:

$$\begin{aligned} HP_1 &\rightarrow w^T x_i + b = +1 \; for \; y_i = +1 \\ HP_2 &\rightarrow w^T x_i + b = -1 \; for \; y_i = -1 \end{aligned} \tag{5}$$

where, $w^T x_i + \varepsilon \geq +1$ gives the hyperplane for the positive class; $w^T x_i + \varepsilon \leq -1$ gives the hyperplane for the negative samples. The two equations in formula (5) can be combined into:

$$y_i(w^T x_i + b) - 1 \geq 0 \quad \forall_i = 1,2,...,\mathrm{n} \tag{6}$$

The SVM margin represents the sum of d_1 and d_2 as:

$$margin = d_1 + d_2 = \frac{2}{||w||} \tag{7}$$

where, d_1 and d_2 are the distance of the samples from the first and second hyperplanes, respectively. In the SVM algorithm, the margin width needs to be maximized as:

$$\begin{aligned} & min \frac{1}{2} ||w||^2 \\ & s.t. \; y_i(w^T x_i + b) - 1 \geq 0 \quad \forall_i = 1,2,...,\mathrm{n} \end{aligned} \tag{8}$$

Combining the objective function ($min \frac{1}{2} ||w||^2$) and the constraint $y_i(w^T x_i + b - 1 \geq 0)$, the binary classification problem can be formalized into the following Lagrange formula:

$$minL_p = \frac{||w||^2}{2} - \sum_I a_i(y_i(w^T x_i + b) - 1) = \frac{||w||^2}{2} - \sum_I a_i(y_i(w^T x_i + b) + \sum_{i=1}^{N} a_i \tag{9}$$

where, a_i is the Lagrange multiplier for x_i; L_p is the primary problem. The values of w, ε, and a that minimize L_p in formula (9) were calculated by differentiating L_p with respect to w and ε and setting the derivatives to zero as:

$$\frac{\partial L_p}{\partial w} = 0 \Rightarrow \mathrm{w} = \sum_{i-1}^{N} \alpha_i y_i x_i \tag{10}$$

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{11}$$

Substituting formulas (10) and (11) into formula (9), the binary classification problem can be transformed as:

$$\begin{aligned} & maxL_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ & s.t. \alpha_i \geq 0, \sum_{i=1}^{N} \alpha_i y_i = 0 \; \forall_i = 1,2,\ldots,N \end{aligned} \tag{12}$$

where, L_D is the dual form of L_p. Then, the values of w, ε, and α can be determined by finding out a solution to formulas (10)~(12). In the SVM, most of the α_i are zeros, an evidence of the sparseness of the algorithm. The nonzero α values are the samples closest to the hyperplane, corresponding to support vectors (SV_s). Hence, the SV_s achieved the maximum width margin.

For nonlinear data, the key of the SVM is to use a nonlinear mapping function, i.e. the kernel function, to map samples into a higher-dimensional linear space. After the mapping, linear classification can be performed in the higher-dimensional space. After kernel transformation by the kernel function $K_{(x,y)}$, the decision function becomes:

$$f_x = \mathrm{w} \cdot \Phi_x + \varepsilon \tag{13}$$

2.6 Optimization

Then, several components of the SVM algorithm was optimized, including the penalty factor C, the kernel function width q and the insensitive band loss function g. The commonly used optimization methods include particle swarm optimization (PSO), and genetic algorithm (GA). After that, the classification effect of our model was evaluated by mean absolute percentage error (MAPE), mean square error (MSE) and Theil's coefficient of inequality.

2.7 Comparative experiment

Finally, the SVM classification performance of the research data were compared with traditional intelligent algorithms like Naive Bayes classifier, logistic regression, J48 tree classifier, and radial basis function (RBF) network. Many scholars have utilized one of these methods to predict information cascades, but few explained the reason to choose a specific method [19-20]. That is why the authors decided to compare the effectiveness of all these methods and verify if the SVM outperform the contrastive methods in predicting the dissemination of health news.

3. RESULTS ANALYSIS

3.1 Ample features

To predict the dissemination trend of a news post, the features whose information gain is above 0.01 were selected and ranked in descending order of information gain (Table 3). As shown in Figure 3, the number of forwards of the news post within 120min, being a social feature, provides the highest information gain (0.6096). This means social features can greatly influence the dissemination of health news. This makes sense in the dissemination of risk information, in that a risk information reposted by numerous users early on can spread to more users in future. Therefore, it is meaningful to include social features of microblog news posts to existing models of information dissemination. Recommendation systems can also utilize social feature like the bandwagon effects [21].

Contrary to a previous study [22], the information gains indicate that author features are crucial to health news dissemination on microblogging sites. The posts published by an active, experienced and authoritative author can propagate widely across the network.

In addition, the content features directly bear on the dissemination trend of microblog news posts. This agrees with the previous research. For example, a title helps to convey the key message of the news, making the post more attractive to users. The proliferation of a public health news post is positively correlated with its comprehensiveness, i.e. the presence of a proper title and the massive use of emoticons, especially neutral emotions.

Table 3. List of the selected features

Number	Name of feature	Information gain
35	Number of forwards of the news post within 120min	0.6096
29	Number of days since the creation of the account	0.1913
32	Length of self-description	0.1408
31	The year of the creation of the account	0.1383
34	Number of new followers attracted by the news post	0.1271
26	Number of users followed by the author	0.0901
25	Number of followers of the author	0.0787
33	Length of verification information	0.0713
24	Number of news posts published since the creation of the account	0.0662
27	Number of news posts favored by the author	0.0448
14	Title length	0.0180
9	Whether the post was published in peak hours?	0.0176
2	Number of neutral emotions	0.0167
30	Gender of the author	0.0153
20	Number of exclamation marks	0.0120
23	Topic length	0.0103

3.2 Optimal classification algorithm

The final classification results of our SVM algorithm are listed in Table 3. For comparison, Naive Bayes classifier, logistic regression, J48 tree classifier, and RBF network were extracted from the Weka library, and also tested on the set of 16 features with high information gain. The classification performance of the SVM and these contrastive algorithms on microblog posts of health news are compared in Table 4. Each method was tested by 10-fold cross-validation. The performance was evaluated pairwise by precision, and F-score.

It can be seen from Table 4 that most of the contrastive methods achieved an average accuracy of no more than 80 %, while the SVM realized the highest Class 1 F-score (0.841), Class 1 precision (0.969), and Class 1 recall (0.769). Therefore, our algorithm outperformed all the contrastive methods in determining whether a news post will be widely disseminated. Although J48 tree classifier achieved relatively high Class 1 precision (0.967), it failed to obtain a sufficiently high Class 1 recall. The research results are consistent with the conclusion in previous studies that the SVM is better than standard machine learning algorithms.

Table 4. Prediction performance of different algorithms

Classification methods	Class 1 precision	Class 1 recall	Class 1 F-score	Weighted average precision
Naive Bayes	0.655	0.789	0.72	0.844
Logistic regression	0.879	0.765	0.818	0.905
J48 tree classifier	0.967	0.731	0.833	0.923
RBF network	0.654	0.748	0.698	0.831
SVM	0.969	0.769	0.841	0.921

4. CONCLUSIONS

Social media like microblogging sites have already developed into 24/7 dissemination platform of real-world events. Despite the research efforts in information dissemination, it remains a huge challenge to monitor or detect the risk information from social media services. For crisis managers, it is extremely important to understand the news propagation on microblogging sites and predict future crisis based on microblog news posts. Therefore, this paper explores deep into the features of microblog posts on health news and their dissemination trends.

This research makes several theoretical contributions. First, the authors identified the inconsistencies of conceptualization in the literature on content research, and re-conceptualized content features according to the persuasion theory. Second, this research incorporates the new realities of microblogging sites into the conceptualization of factors related to attitude strength. Third, this research highlights the importance of social information and the key role of author features in news evaluation.

There are also direct practical implications of our research. For example, the authors developed a set of ample, diverse features and an optimal algorithm to predict news dissemination. These results provide a valuable reference to researchers on information cascades. Furthermore, our approach targets specifically the news on public health issues. Thus, our analysis informs crisis managers about public reaction towards specific health news. Finally, our method is computationally feasible in near real-time scenarios and can be utilized to capture rapidly changing dynamics of microblog news dissemination.

ACKNOWLEDGEMENT

This paper is supported by National Natural Science Foundation of China (Grant No.: 71871106), Jiangsu Provincial Social Science Foundation (Grant No.: 18GLD014), Jiangsu Provincial University Philosophy and Social Science Foundation (Grant No.: 2018SJA0815), the Fundamental Research Funds for the Central Universities (Grant No.: 2019JDZD06) and Chinese University Research Foundation for Young Scholars (Grant No.: JUSRP11882).

REFERENCES

[1] Ucar, T., Culpan, M., Caskurlu, T., Karaman, M., Silay, M.S. (2018). The activity and discussion points of #Circumcision through Twitter; a microblogging platform. International Journal of Impotence Research, 30(5): 249-252. http://dx.doi.org/10.1038/s41443-018-0058-y

[2] Hasan, M., Orgun, M.A., Schwitter, R. (2019). Real-time event detection from the Twitter data stream using the TwitterNews+ Framework. Information Processing & Management, 56(3): 1146-1165. http://dx.doi.org/10.1016/j.ipm.2018.03.001

[3] Akcura, T., Altinkemer, K., Chen, H.L. (2018). Noninfluentials and information dissemination in the microblogging community. Information Technology & Management, 19(2): 89-106. http://dx.doi.org/10.1007/s10799-017-0274-z

[4] Carrera, B., Jung, J.Y. (2018). SentiFlow: An information diffusion process discovery based on topic and sentiment from online social networks. Sustainability, 10(8): 2731. http://dx.doi.org/10.3390/su10082731

[5] Gursoy, D. (2019). A critical review of determinants of information search behavior and utilization of online reviews in decision making process (invited paper for 'luminaries' special issue of International Journal of Hospitality Management). International Journal of Hospitality Management, 76: 53-60. http://dx.doi.org/https://doi.org/10.1016/j.ijhm.2018.060 03

[6] Pei, J.Y., Yu, G., Tian, X.Y., Donnelley, M.R. (2017). A new method for early detection of mass concern about public health issues. Journal of Risk Research, 20(4): 516-532. http://dx.doi.org/10.1080/13669877.2015.1100655

[7] Wu, L.R., Li, J.J., Qi, J.Y. (2019). Modeling information popularity dynamics based on branching process. Acta Physica Sinica, 68(7): 6. http://dx.doi.org/10.7498/aps.68.20181948

[8] Yan, J., Zhou, Y., Wang, S.Y., Li, J. (2019). To share or not to Share? Credibility and dissemination of electric vehicle-related information on WeChat: A moderated dual-process model. IEEE Access, 7: 46808-46821. http://dx.doi.org/10.1109/ACCESS.2019.2909072

[9] Aroean, L., Dousios, D., Michaelidou, N. (2019). Exploring interaction differences in Microblogging Word of Mouth between entrepreneurial and conventional service providers. Computers in Human Behavior, 95: 324-336. http://dx.doi.org/10.1016/j.chb.2018.10.020

[10] Cheung, C.M.K., Thadani, D.R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. Decision Support Systems, 54(1): 461-470. http://dx.doi.org/10.1016/j.dss.2012.06.008

[11] Zhang, L., Peng, T.Q., Zhang, Y.P., Wang, X.H., Zhu, J.J.H. (2014). Content or context: Which matters more in information processing on microblogging sites. Computers in Human Behavior, 31: 242-249. http://dx.doi.org/10.1016/j.chb.2013.10.031

[12] Pei, J., Yu, G., Shan, P. (2016). Social media coverage of public health issues in China: A content analysis of Weibo news posts. Information Technology: New Generations, 448: 111-120. https://doi.org/10.1007/978-3-319-32467-8_11

[13] Nerlich, B., Forsyth, R., Clarke, D. (2012). Climate in the news: How differences in media discourse between the US and UK reflect national priorities. Environmental Communication: A Journal of Nature and Culture, 6(1): 44-63. http://dx.doi.org/10.1080/17524032.2011.644633

[14] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. (2009). The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, 11(1): 10-18. http://dx.doi.org/10.1145/1656274.1656278

[15] Shannon, C.E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3): 379-423. http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x

[16] Colombo, E., Forte, G., Rossignoli, R. (2019). Carry trade returns with support vector machines. International Review of Finance, 19(3): 483-504. http://dx.doi.org/10.1111/irfi.12186

[17] Kumar, B., Vyas, O.P., Vyas, R. (2019). A comprehensive review on the variants of support vector machines. Modern Physics Letters B, 33(25): 1950303. http://dx.doi.org/Artn195030310.1142/S021798491950 3032

[18] Sowah, R.A., Kuuboore, M., Ofoli, A., Kwofie, S., Asiedu, L., Koumadi, K.M., Apeadu, K.O. (2019). Decision support system (DSS) for fraud detection in health insurance claims using genetic support vector machines (GSVMs). Journal of Engineering, 2019:

1432597. http://dx.doi.org/Artn 143259710.1155/2019/1432597
[19] Orkphol, K., Yang, W. (2019). Sentiment analysis on microblogging with K-means clustering and artificial bee colony. International Journal of Computational Intelligence and Applications, 18(3): 1950017. http://dx.doi.org/Artn195001710.1142/S1469026819500172
[20] Vazquez, S., Munoz-Garcia, O., Campanella, I., Poch, M., Fisas, B., Bel, N., Andreu, G. (2014). A classification of user-generated content into consumer decision journey stages. Neural Netw, 58: 68-81. http://dx.doi.org/10.1016/j.neunet.2014.05.026
[21] Choi, S.M., Lee, H., Han, Y.S., Man, K.L., Chong, W.K. (2015). A recommendation model using the bandwagon effect for E-marketing purposes in IoT. International Journal of Distributed Sensor Networks, 2015: 1-7. http://dx.doi.org/10.1155/2015/475163
[22] Masip, P., Guallar, J., Suau, J., Ruiz-Caballero, C., Peralta, M. (2015). News and social networks: audience behavior. El Profesional de la Informacion, 24(4): 363. http://dx.doi.org/10.3145/epi.2015.jul.02