

Automatic Foot-and-Mouth Disease (FMD) News Monitoring and Classification

Yulei Zhang¹, Chunju Tseng¹, Yan Dang¹, Hsinchun Chen¹, Catherine A. Larson¹, Zack Whedbee², Andres Perez², Mark Thurmond²

¹Artificial Intelligence Lab, Department of Management Information Systems, Eller College of Management, University of Arizona, Tucson, AZ 85721, USA; ²FMD Lab, Center for Animal Disease Modeling and Surveillance (CADMS), University of California, Davis, CA 95616

OBJECTIVE

This paper describes an FMD news monitoring and classification system which can automatically monitor FMD related news from online news data sources, generate news summarization and classify the news into three categories defined by domain experts. The report research is a collaborative effort between the Artificial Intelligence Lab at the University of Arizona and the FMD Lab at UC Davis.

BACKGROUND

Foot-and-mouth disease (FMD) is one of the most devastating diseases of farm animals. There is a critical need for countries to have a global FMD situational awareness. Monitoring the online news sources for FMD-related news is an important component of situational awareness. The FMD Lab at UC Davis (<http://fmd.ucdavis.edu/>) has developed models and systems for global FMD surveillance, including the FMD BioPortal web-based system jointly with the AI Lab at the University of Arizona. They have also been gathering and processing FMD-related news from the FMD World Reference Laboratory, the OIE, the FAO, among others. However, manual searches are necessary identify and integrate the FMD news into their models and systems. This manual work is not only time consuming and labor intensive but may also lead to the loss of some important information.

It is difficult for general-purpose search engines to keep up with the increasing size of the Web [1]. However, domain-specific online news collection and analysis has been successfully used to get timely and useful information. The dark web project [2] collected terrorism information from the Web and did Web community visualization of domestic extremists and hate groups. The CMedPort [3] was developed to facilitate online medical information seeking in the medical domain. Pant created a topical spider to get a document collection which can help to locate relevant business entities [4].

METHODS

The main components of the FMD News Portal system framework include data acquisition, summarization and classification. We have set up a spidering program to regularly gather the FMD-related news from the Internet. After spidering, a

key-words-based filtering program filters out the FMD news from the large collection of documents retrieved by our spiders. Then the news articles are stored into a relational database. An automatic summarization tool is employed to extract the main points of these news articles and improve the performance of the classifiers. The classification component uses machine learning algorithms to automatically classify the news, which can be either generated summaries or raw news articles, into three categories suggested by the domain experts in the FMD Lab. The three categories currently being processed by our system are: (1) FMD outbreak related news, (2) FMD control program related news, and (3) FMD social, economic and general information.

RESULTS

From the thirty three English websites being monitored at present, we have spidered 183,241 items. In the classification experiments, we used Bag of Words, Noun Phrases and Name Entities features to represent the news. We compared different machine learning algorithms for classification and achieved the highest precision of 75.93%.

CONCLUSIONS

The automatic FMD news monitoring system is useful for FMD surveillance. With such a system, we can automatically gather FMD related news from the Web and organize the information into different categories. This approach can also be extended to monitor other diseases. Future research will address bilingual processing through a machine translation component.

REFERENCES

- [1] M. Chau and H. Chen, "Comparison of Three Vertical Search Spiders," *Computer*, 36:5, May 2003, pp 56-62.
- [2] Zhou, Y., Reid, E., Qin, J., Chen, H. and Lai, G., "US Domestic Extremist Groups on the Web: Link and Content Analysis" *IEEE Intelligent Systems*, 20:5, 2005, pp. 44-51.
- [3] Zhou, Y., Qin, J., and Chen, H., "CMedPort: An integrated approach to facilitating Chinese medical information seeking," *Decision Support Systems*, 42:3, December 2006, pp. 1431-1448.
- [4] Pant, G. and Menczer, F. 2003. "Topical crawling for business intelligence." In *Proceedings of the 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)*. Trondheim, Norway.