

# A Novel Knowledge Network Framework for Financial News Navigation

Lili Zhou<sup>1</sup>, Hanchao Wang<sup>1</sup>, Lei Zhang<sup>1</sup>, Enhong Chen<sup>1</sup>, and Jun Chen<sup>2</sup>

<sup>1</sup> University of Science and Technology of China  
{zhoulili,hanchaow,stone}@mail.ustc.edu.cn

<sup>2</sup> Xinhua News Agency  
cheneh@ustc.edu.cn, chenjun2008@xinhua.org

**Abstract.** Nowadays, various financial news retrieval platforms are provided to help users, especially for financial professionals and hobbyists to make right decisions. In those platforms, users usually get information by searching the relevant news via keywords or clicking the recommended news with the similar topic in the clicked web page. However, such ways to obtain financial information cannot effectively meet users' further needs. They are eager to obtain the relevant news with different domains in a short time. To address this problem, we propose a novel four-layers-based knowledge network framework for financial news navigation. Experiments on real data sets demonstrate the effectiveness and efficiency of our proposed framework.

**Keywords:** retrieval platform, knowledge network, financial news navigation.

## 1 Introduction

Currently, lots of financial news retrieval platforms including general search engines (e.g., baidu) and financial domain vertical websites (e.g., sina) arise to facilitate access to financial information for users, especially for financial professionals and hobbyists. They are eager to get the latest financial information to help them make right decisions. In these platforms, users can search the relevant news by keywords or click the corresponding label in the navigation menu to obtain financial information they need. If wanting to know the ins and outs of the clicked news, users have to continue to seek those related news through the aforementioned two ways. Such a process is time-consuming and laborious. Although the majority of news pages list some recommended news links with the similar topic for extension reading, this does not alleviate the problem. Meanwhile, existing studies on financial news mainly focus on how to organize and present financial news for users more friendly [1] or mining the underlying information in financial news [2]. But All of them do not solve this problem properly.

For this, we propose a novel four-layers-based knowledge network framework for financial news navigation. Specifically, we first crawl large amounts of financial news from many popular financial web portals. Second, we apply topic model

and classification methods to extract knowledge from the collected news corpus. Based on these knowledge sets, we then construct a four-layers-based knowledge subnet for each news report. Finally, we implement and visualize this knowledge network with a popular used library D3.js.

## 2 The Four-Layers-Based Knowledge Network

Constructing the knowledge network framework consists of two steps, i.e., offline step and online step. Next, we detail the involved techniques and implementations respectively.

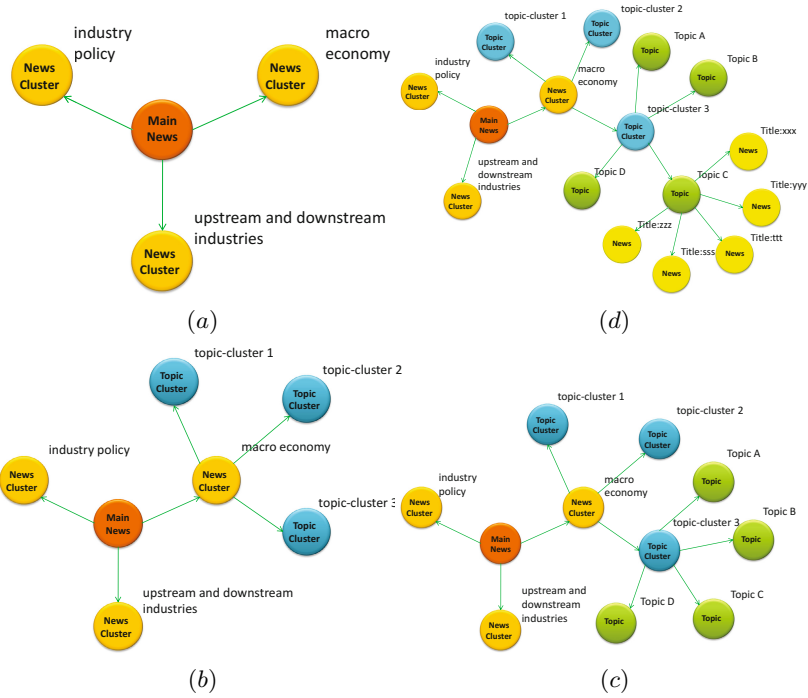
### 2.1 Knowledge Definition and Extraction

In the offline step, we crawl real datasets and extract the knowledge from them. Here we define knowledge as financial information contained in financial news, which can be expressed as an industry label or a topic, etc. Among these knowledge, the topics and industry label of one news report are usually difficult to obtain. Fortunately, Latent Dirichlet Allocation [3] which is a classic topic model can be used to extract the latent topics from vast amounts of financial news effectively. Since the headline often represents the core idea of a news report, we consider both the title and news content and give different weights to them when using LDA model. To recognize the industry label, we choose Support Vector Machine [4] due to its high accuracy and efficiency.

### 2.2 Knowledge Network Construction and Visualization

In the online step, we construct the dynamic knowledge network composed of numerous knowledge subnets. Fig.1 shows the full view.

Suppose a user opens a news page, then the first layer of the knowledge subnet for this news report shown in Fig.1(a) is presented behind the news content. In Fig.1(a), the center node denotes the clicked news (hereafter we call Main News) and each of the other three nodes connecting to Main News stands for a news set of the common nature. For example, if Main News's industry label is real estate, the node named "industry policy" represents some relevant news whose labels are all real estate control policy. If the user is interested in macro data after reading Main News, he or she can click the corresponding news-cluster node to get detailed information. Fig.1(b) shows the second layer of the knowledge subnet where each blue node denotes a similar topic set with a description of several words and each topic in the set is contained by some news in its parent node. We use a simple clustering method named K-Means [5] to aggregate those similar topics. And click one blue node, the specific topics in this set are displayed shown in Fig.1(c). Furthermore, the user also can find top-n relevant news for one interesting topic by clicking the corresponding topic-node. As shown in Fig.1(d), there are 5 macro economy news reports whose main topics are all "Topic C". We find the top-5 news reports according to the probability value of this topic



**Fig. 1.** The structure of a knowledge subnet for one financial news report

in all news reports. Finally, the user can read one of these news by clicking it. Through the whole process, users can conveniently grasp the information with different topics

We visualize each layer of the knowledge network with a popular used library D3.js (<http://d3js.org/>) which can bring data to life using HTML, SVG and CSS.

### 3 Experiments

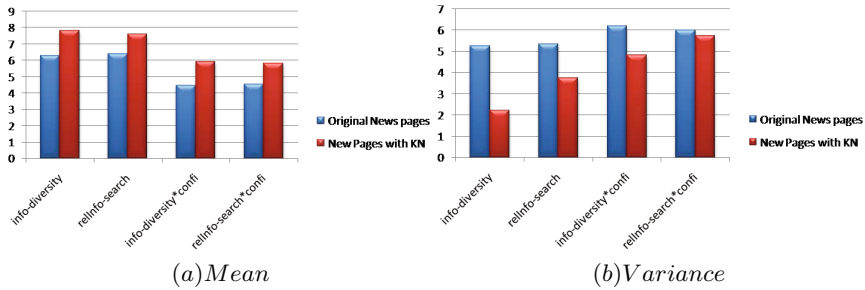
To validate the proposed framework, we implement our proposed framework focusing on real estate according to Section 2. All the implementations in Java are on a Windows 7 PC with Intel 1-core i3 3.10GHz CPU, 4GB of main memory and a 64-bit operating system. On this basis, we design a scoring system for comparing two different kinds of news presentation, that is, the novel news page with the knowledge subnet embedded in denoted as Page 1 and the original news page denoted as Page 2. We offer four measures to users for marking the two different pages. Table 1 presents meanings of these measures. In order to avoid human bias, we deploy this scoring system on the server to let external users grade them after using it.

So far, there are more than two thousand scoring records. We randomly select 600 pieces of data among them for final statistics. Among these users, 399 users

**Table 1.** Details of four scoring indexes for two financial news presentations

| Measure         | Annotation   |
|-----------------|--|
| info-diversity  | information diversity of extended news sets (1-10')                                |
| relInfo-search  | whether help users find related news and offer a positive user experience. (1-10') |
| user-confidence | users' confidence for the scores they give (1-10')                                 |
| whether-like    | whether like such a way of news presentation (0/1)                                 |

express a preference for Page 1. As shown in Fig.2(a), the average scores of Page 1 are both higher than which of Page 2 for info-diversity and relInfo-search, similar to the other situation that we combine users' confidence for their ratings. Higher users' confidence means that the results they graded are more credible. Meanwhile, the variances of scores of Page 1 are lower than that of Page 2 for the two indexes in both cases. In addition, we apply z-test and find that the differences between the ratings obtained by our proposed approach and the exiting news presentations (e.g., sina) are statistically significant with  $|z| \geq 2.58$  and thus  $p \leq 0.01$ . Therefore, our proposed framework for financial news navigation outperforms the existing other news presentations.

**Fig. 2.** The scoring results: mean and variance

## 4 Conclusion

In this study, we presented a four-layers-based knowledge network framework for financial news navigation. The experiments focusing on real estate industry demonstrated that the proposed framework can effectively satisfy users' further needs. Furthermore, the idea of using knowledge network to facilitate users' access to financial information can be generally applicable to other domains of news navigation (e.g., sports news). In the future, we would like to incorporate the behavior data such as geographic information and users' browsing history logs into the proposed framework for personalized recommendation.

**Acknowledgments.** This research was partially supported by grants from the Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20113402110024), the Science and Technology Development of Anhui

Province, China (Grants No. 13Z02008-5 and 1301022064), the International Science & Technology Cooperation Plan of Anhui Province (Grant No. 1303063008), the Nature Science Foundation of Anhui Education Department (Grant No. KJ2012A273).

## References

1. Wang, H., Wang, Z.: Mobile financial news mashup development based on yql. In: 2013 Fifth International Conference on Computational and Information Sciences (ICCIS), pp. 1717–1720. IEEE (2013)
2. Alanyali, M., Moat, H.S., Preis, T.: Quantifying the relationship between financial news and the stock market. *Scientific Reports* 3 (2013)
3. Blei, D.M., Ng, A.Y., Jordan, M.: Latent dirichlet allocation. *The Journal of Machine Learning Research* 3, 993–1022 (2003)
4. Sebastiani, F.: Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)* 34(1), 1–47 (2002)
5. Jain, A.K.: Data clustering: 50 years beyond k-means. *Pattern Recognition Letters* 31(8), 651–666 (2010)