Visualizing Technology Trends

Zhenhuan Chen, Yong Bai, Jie Tang, Yutao Zhang, Bo Gao, and Debing Liu
Department of Computer Science and Technology, Tsinghua University
{chenzhh3671,zhangyutao1106,elivoa,dbliu384}@gmail.com,jietang@tsinghua.edu.cn

ABSTRACT

In this paper, we introduce an interesting function, Trend[1], in AMiner [4]. Given a technology field, Trend can automatically extract sub-topics, visualize, and compare their evolutionary trends.It can also identify "milestones" happened in the technology field. The system is publicly available. Feedback from online users shows the effectiveness of the system.

CCS CONCEPTS

• **Information systems → Data mining**; • **Human-centered computing → Visualization application domains**.

KEYWORDS

Visualization, Trend Analysis, Data Mining

ACM Reference Format:
Zhenhuan Chen, Yong Bai, Jie Tang, Yutao Zhang, Bo Gao, and Debing Liu . 2019. Visualizing Technology Trends . In *Proceedings of Anchorage '19: ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (Anchorage '19).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Technology trends can provide the user with a concise summary of each field, gain insights on how it evolves over time, and what are the milestones happened during the development of the field. Analysis of technology trend can help both research and industry communities understand the history and the current situation, and grasp the frontier of a field. We have developed a trend visualization system, referred to as Trend, with the goal of visualizing and comparing the evolutionary patterns of different topics in a research field. Trendis based on the literature text information from AMiner [4], which consists of more than 130,000,000 researchers and 270,000,000 publication papers. AMiner is an academic search and mining system, supporting expert finding [3], rising star ranking, collaborator recommendation [2], reviewer suggestion [1], name disambiguation [5], and so on.

To build the trend visualization system, technical challenges lie in: 1) *sub-topics* – what are the major topics in a research field? 2) *visualization* – how to visualize and compare the evolutionary patterns of different topics; and 3) *explainability* – how to explain the visualized evolutionary patterns.

In Trend, we use a two-step method to extract the major topics for a research field. For each topic, we use the frequency of topics to quantify its popularity over time. To enhance the explainability, we identify the milestone events (the most important publications during a period) to explain the visualized topic trend. The Trend system has been put into operation since 2018 and feedback from users demonstrates the effectiveness of the system.

2 KEY FEATURES

Figure 1 shows the screenshot of the Trend system for the research field "Data Mining". On the left side, the system automatically extracts the major topics, e.g., "association rule", "knowledge discovery", "decision tree". For each topic, the strength of the blue bar indicates its popularity. In the middle, the screenshot shows the evolutionary pattern of each topic over time. For each time stamp, if a topic is located at the topic, then the topic is said to be the most popular topic around that time. For example, the orange stream, representing "association rule" was very popular in 1990s, and slowly becoming silence in these years. Instead, "feature extraction", "neural network", and "machine learning" are becoming more and more popular. In each topic-based stream, Trend identifies the most important publications as the miletones. For example, here, the paper entitled "efficient and effective clustering methods for spatial data mining" has been identified as a milestone work. In the rest of this section, we describe with a bit more details about how we implement the Trend system.

Topic Extraction. It is non-trivial to extract the major topics. Too general topics would result in limited useful information to the user, while too specific topics may not easy to cover the field. We finally use a two-step idea to deal with this problem. We first use AMiner [4] to retrieve top experts for the queried field. For example, for "data mining", the system would return "Jiawei Han", "Christos Faloutsos", "Philip Yu", "Vipin Kumar", and so on. Then we extract their research interests and merge those interests together. The highly frequently occurred interests are then considered as the major research topics in this field.

Topic Streamlines. For each extracted topic, we use the number of papers published on it in each year to indicate its popularity over time. Accordingly, topic-based streamlines has been generated for all extracted topics. The streamlines can be used to easily compare different topics and also the rise-and-fall patterns of topics. According to the published year of these literature, a series of time zones are divided, and the literature within one zone will affect the field popularity of that zone. The widths of each streamline chart represent the popularity, and the relative positions of the streamline chart represent the popularity ranking of the term in this zone. The

[1]https://trend.aminer.org

Anchorage '19, August 04–08, 2019, Anchorage, AK

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

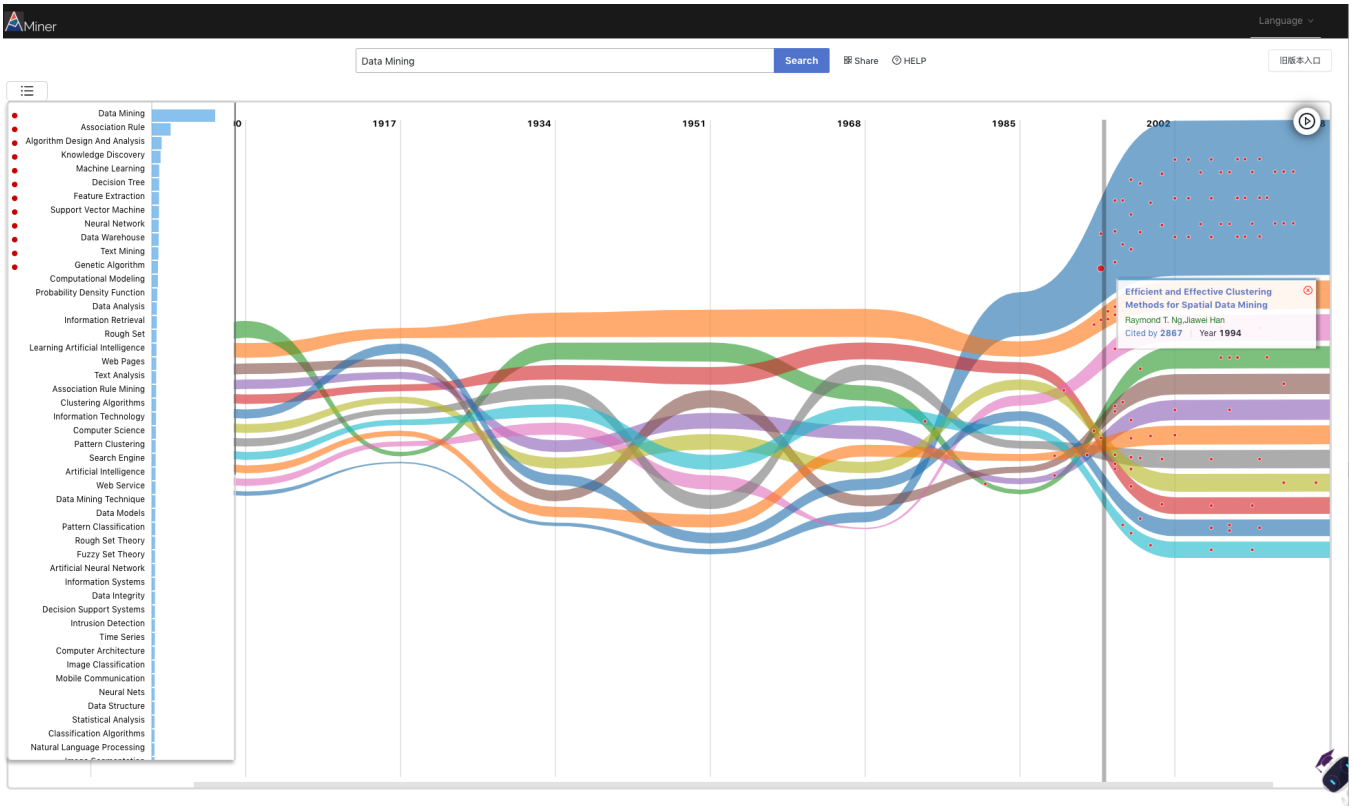

Figure 1: An overview of Trend Analysis System

change of positions between different streamline charts represents the changing trend of popularity in related fields over time.

Topic Milestones. Identifying the most important milestones can be used to help explain the visualized streamlines. In Trend, this has been done simply using the number of citations, but with diversity. Specifically, we first retrieve related papers with the highest number of citations. To guarantee the diversity, we add a penalty factor to those selected papers on the same topic. Finally, the selected milestone papers represent the highly cited research work and also cover many different topics. To visualize the milestones, when the mouse hovers over an event node, a floating box will pop up to show the title, year, citation number and author information of relevant documents. By clicking on the title, you can jump to the corresponding page of AMiner to view more detailed information, which is convenient for users to dive into the results.

Finally, in order to display and compare these important academic achievements as a whole, we have provided a "dynamic timeline" function. When dragging the gray dynamic timeline or clicking on an area of a streamline, a side box will pop up on the left side to show the important events represented by all event nodes in the current year and their related information. By comparing and looking at the important events in different fields in the same year, the relationship among them can be analyzed conveniently, helping users to explore the possibility of mutual influence and cross cooperation between different fields. By clicking the suspended play button, the dynamic timeline will move forward from the current position to dynamically display the event information of the passed year.

3 CONCLUSIONS

The Trend system is publicly available. We use the system to study different research fields and compare the trends of publication venues. The visualization of the discovered trends can significantly help the user to understand technology trends and their evolutionary patterns. As the future work, we are going to work on new methodologies to predict the technique trend and to discover hidden patterns in order to explain the visualized trends.

Acknolwedgements. The work is supported by NSFC (61825602).

REFERENCES

[1] Yujie Qian, Jie Tang, and Kan Wu. 2018. Weakly Learning to Match Experts in Online Community. In *IJCAI'18*. 3841–3847.
[2] Jie Tang, Sen Wu, Jimeng Sun, and Hang Su. 2012. Cross-domain Collaboration Recommendation. In *KDD'12*. 1285–1294.
[3] Jie Tang, Jing Zhang, Ruoming Jin, Zi Yang, Keke Cai, Li Zhang, and Zhong Su. 2011. Topic Level Expertise Search over Heterogeneous Networks. *Machine Learning Journal* 82, 2 (2011), 211–237.
[4] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. ArnetMiner: Extraction and Mining of Academic Social Networks. In *KDD'08*. 990–998.
[5] Yutao Zhang, Fanjin Zhang, Peiran Yao, and Jie Tang. 2018. Name Disambiguation in AMiner: Clustering, Maintenance, and Human in the Loop.. In *KDD'18*. 1002–1011.