Literature Survey

1.On-line new event detection and tracking

By James Allan,Ron Papka and Victor Lavrenko

This paper defines and describes the related problem of new event detection and event tracking within a stream of broadcast news stories. The approach taken in this paper uses a single pass clustering algorithm and a novel thresholding model that incorporates the properties of events as a major component.New event detection and event tracking are a part of Topic detection and tracking (TDT) initiative.

New event detection requires identifying those new stories that discuss an event that has not already been reported in an earlier story.

Event tracking means starting from a few sample stories and finding all the subsequent event that discuss the same story.

2.Dynamic Incremental Analysis of Sub-Topic Evolution

By:Li Fenghuan, Zheng Dequan, Zhao Tiejun

There has been increasing interest in the follow-up progress of events because of sustainability and mutual influence of events. Meanwhile, more and more emergent events make it necessary to follow events in an intuitive and efficient way. However, the majority of traditional event analysis is sentence-oriented or topic-oriented which is event extraction or topic detection and tracking. A hierarchical structure of the topic event is constructed according to the research objects and the scope. A dynamic incremental model is presented for analyzing sub-topic dynamics in the topic event, which borrows the ideas of single-pass clustering, multi-category and dynamic incremental model. It is document-oriented and built on temporal property of the topic event, including dynamic threshold selection, similarity smoothing and dynamic incremental strategy. Meanwhile, overall evaluation criteria combined with χ\+2-test is served for performance analysis. The algorithm is effective and facilitates users to follow the topic event explicitly. Experimental results reported for four well-known topic events in China show that the performance of sub-topic evolution analysis is improved significantly.

3.Metro maps of information

By Dafna Shahaf, Carlos Guestrin and Eric Horvitz

When information is abundant, it becomes increasingly difficult to fit nuggets of knowledge into a single coherent picture. Complex stories spaghetti into branches, side stories, and intertwining narratives. In order to explore these stories, one needs a map to navigate unfamiliar territory.In this they have developed a methodology for creating structured summaries of information, which we call metro maps. Their algorithm generates a concise structured set of documents which maximizes coverage of salient pieces of information. Most importantly, metro maps explicitly show the relations among retrieved pieces in a way that captures the evolution of a story. They first formalized characteristics of good maps and formulate their construction as an optimization problem. Then, they provided efficient methods with theoretical guarantees for generating maps. Finally,they integrated capabilities for supporting user interaction into the framework, allowing users to guide the formulation of the maps so as to better re ect their interests. Pilot user studies with a real- world dataset demonstrate that the method is able to produce maps which help users to acquire knowledge efficiently.

4.Text Clustering Incremental Algorithm in Sensitive Topic Detection

By Yuejin Zhang , Jiajia Zhang, Dongmei Zhao

With the rapid development of Internet technology, the influence of online consensus continues to expand. How to quickly and effectively discover sensitive topics and keep track of those topics has become an important research recently. Text clustering can aggregate news texts with the same or similar content to achieve the purpose of discovering topics automatically. Make improvement to clustering algorithm according to different media types is the main research direction. Although the existing typical clustering algorithms have certain advantages, they all face constraints on data size and data characteristics in specific applications. There is no existing algorithm can fully adapt to these characteristics. Although the application of more Single-pass algorithms in the (TDT) field can realize the discovery and tracking of topics, there are disadvantages of poor accuracy and slow speed under massive data. According to the dynamic evolution characteristics of online consensus, this paper proposes an incremental text clustering algorithm based on Single-pass, which optimizes the clustering accuracy and efficiency of massive news. Based on the real online news texts from the online consensus analysis system, they conducted an experiment to test and verify the feasibility and effectiveness of the algorithm we proposed. The result shows that the new algorithm is much more efficient compared to the original Single-pass clustering algorithm. In the real application, the new incremental text clustering algorithm basically meets the real-time demand of online topic detection and has a certain practical value.