Metro Maps of Information

Dafna Shahaf
Stanford University
and
Carlos Guestrin
University of Washington
and
Eric Horvitz
Microsoft Reseach

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. We have developed a methodology for creating structured summaries of information, which we call *metro maps*. Our 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. We first formalize characteristics of good maps and formulate their construction as an optimization problem. Then, we provide efficient methods with theoretical guarantees for generating maps. Finally, we integrate capabilities for supporting user interaction into the framework, allowing users to guide the formulation of the maps so as to better reflect 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.

1. INTRODUCTION

"Distringit librorum multitudo" (the abundance of books is a distraction), said Lucius Annaeus Seneca; he lived in the first century.

A lot has changed since the first century, but Seneca's problem has only become worse. The surge of the Web has brought down the barriers of distribution, and people find themselves overwhelmed by the increasing amounts of data. In the midst of a glut of information, relevant information is often buried in an avalanche of data, and locating it is difficult. The problem spans entire sectors, from intelligence analysts to scientists and web users.

In recent years, search engines have been relied upon for accessing information, and specialized tools were developed (e.g., academic search and news search). However, while search engines are highly effective in retrieving nuggets of knowledge, the task of fitting those nuggets into a coherent picture remains difficult.

Several methods have attempted to summarize and visualize complex stories [Swan and

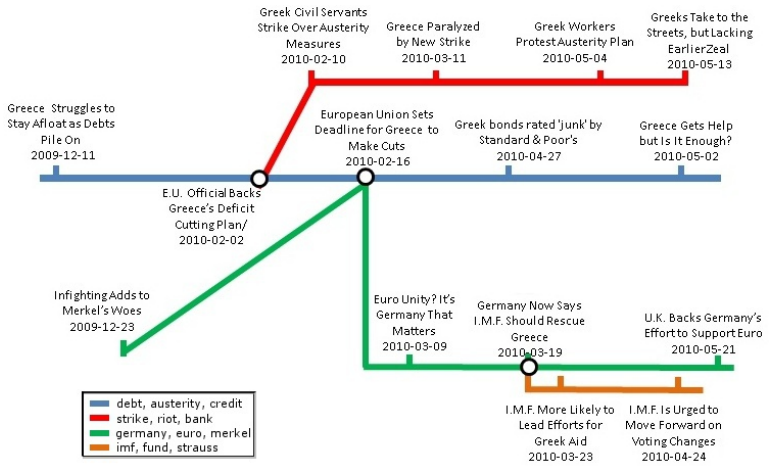

Fig. 1. A sample metro map, computed for the query 'Gree* debt'. The main storylines discuss the austerity plans, the riots, and the role of Germany and the IMF in the crisis.

Jensen 2000; Yan et al. 2011; Allan et al. 2001]. However, most of these methods work only for simple stories, which are linear in nature. In contrast, complex stories exhibit a very non-linear structure: stories spaghetti into branches, side stories, dead ends, and intertwining narratives. To explore these stories, users needs a *map* to guide them through unfamiliar territory.

In this paper we summarize methods we have developed for automatically creating *metro maps* of information [Shahaf et al. 2012b; 2012a]. Metro maps consist of a set of lines which have intersections or overlaps. Lines follow coherent narrative threads; different lines focus on different aspects of the story. We found that this visualization allows users to easily digest information at a holistic level, interact with the model and make modifications.

We show an example metro map in Figure 1 for the query 'Gree* debt'. The main storylines discuss the austerity plan, the riots, and the role of Germany and the IMF. Note how the blue and red lines intersect at an article about the austerity plan, as the plan plays an important role in both storylines: it was a key precondition for Greece's bailout, and also triggered many of the strikes. We believe that metro maps can serve as effective tools to help users cope with information overload in many fields, and frames a direction for research on the automated extraction of information and construction of new representations for summarizing and presenting complex sets of interrelated concepts.

2. CRAFTING AN OBJECTIVE FUNCTION

The problem of finding a good metro map is hard, especially because it is not clear what we are looking for. Recognizing whether a map is good or not is easy for humans, but it is a very intuitive notion. In the following, we review desired properties of a metro map. We motivate and formalize several (sometimes conflicting) criteria.

Coherence. A first requirement is that each line tells a *coherent* story: Following the articles along a line should give the user a clear understanding of the evolution of a story.

A natural first step in defining coherence is to measure similarity between each pair of consecutive documents along the line. However, local connections may give rise to associative, incoherent lines. Figure 2 demonstrates this idea: Suppose a line consists of three documents, each containing four words. Documents 1 and 2 are similar, and so are 2 and 3. However, documents 1 and 3 have nothing in common.

Fig. 2. A line of length 3. Each pair of consecutive documents is similar, but the line is not coherent.

Coherent lines, on the other hand, can often be characterized by a *small* set of words, which are important throughout many of the transitions. In other words, coherence is a *global* property of the line, and cannot be deduced solely from local interactions.

We transform the problem into an optimization problem, where the goal is to choose a small set of words, and score the line $d_1, ..., d_n$ based only on these words. In order to ensure that each transition is strong, the score of a line (given a set of active words) is the score of the weakest link.

$$\textit{Coherence}(d_1, ..., d_n) = \max_{W \in \textit{activations}} \textit{Coherence}(d_1, .., d_n|W)$$

$$\textit{Coherence}(d_1, .., d_n|W) = \min_{i=1...n-1} \textit{score}(d_i \rightarrow d_{i+1}|W)$$

The way to score a link might depend on the domain: In [Shahaf et al. 2012b] we show how to compute a score given article content alone. In [Shahaf et al. 2012a], we show how to take advantage of meta-data, in the form of a citation graph, to compute a score that takes scientific influence into account.

Coverage. *Coherence* is crucial for good maps, but is it sufficient? In order to pursue an answer to this question, we found maximally-coherent lines for the query 'Bill Clinton'. The results were discouraging. While the lines we found were indeed coherent, they were not *important*. Many of the lines revolved around narrow topics, such as Clinton's visit to Belfast. Furthermore, as there was no notion of diversity, multiple lines contained redundant information. This example suggests that selecting the most coherent lines does not guarantee a good map. Instead, the key challenge is balancing coherence and *coverage*: in addition to being coherent, lines should also cover diverse topics which are important to the user.

We define a set of elements that we wish to cover, E. These elements can depend on the domain: In the case of news articles, we chose words ("Obama", "China") [Shahaf et al. 2012b]; in the case of a scientific corpus, we chose papers [Shahaf et al. 2012a], so a high-coverage map has impact on a large chunk of the corpus.

We calculate a function $\textit{cover}_d(e)$, measuring how well document d covers element e. We then extend $\textit{cover}_d(e)$ to functions measuring the coverage of sets of documents, $\textit{cover}_D(e)$. In order to encourage diversity, this function should be *submodular*. Thus, if the map already covers e well, adding another document which covers e well provides very little extra coverage. This encourages us to pick documents that cover new topics.

Next, we introduce weights for each element in E, and set map coverage to be $\sum_e \lambda_e \cdot \textit{cover}_D(e)$. The weights bias the map towards covering important elements, while also offering a natural mechanism for personalization. In [Shahaf et al. 2012b], we discuss

learning weights from user feedback, resulting in a personalized notion of coverage.

Connectivity. Finally, a map is more than just a set of lines; there is information in its *structure* as well. Therefore, our last property is *connectivity*. The map should convey the underlying structure of the story, and how different aspects of the story interact with each other. In order to encourage intersections, we simply define connectivity as the number of line pairs that intersect: $\sum_{i<j} \mathbb{1}(line_i \cap line_j \neq \emptyset)$.

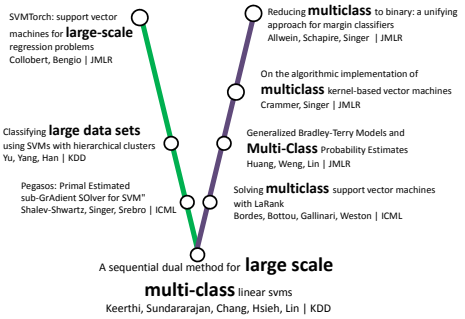

Fig. 3. A fragment of a science map for the query SVM, showing intersection of two lines: multi-class SVMs and large-scale SVMs.

3. ALGORITHM

We now briefly review the main ideas behind the algorithm. First, in order to pick coherent lines, we wish to identify all possible candidates. However, representing all coherent lines is infeasible. Instead we propose a divide-and-conquer approach, constructing long lines from shorter ones. This approach allows us to compactly encode many candidate lines as a graph; the nodes of the graph correspond to short coherent lines, and edges indicate lines that can be concatenated and remain coherent.

Next, we use the graph to find a set of K coherent lines that maximize coverage. Finding these lines is an NP-hard problem, but we take advantage of the submodularity of $cover_D(\cdot)$ and apply an approximation algorithm with theoretical guarantees. Finally, we perform local search, trying to substitute lines in a way that increases connectivity without sacrificing coverage. Example outputs of the algorithm are depicted in Figures 1 and 3.

4. EVALUATION

We evaluated the effectiveness of metro maps in aiding users navigate, consume, and integrate different aspects of a story. We describe two such studies below:

Scientific Domain. In the first study, we recruited graduate students and asked them to conduct a quick literature survey in an area they were not familiar with. Participants used Google Scholar [Goo], a search engine that indexes scholarly literature. Half of the participants were also given a metro map.

An expert graded the participants' output. We measured *precision* (fraction of retrieved papers that are relevant), and *recall* (fraction of relevant research areas retrieved). It is not enough for the users to find good papers; rather, it is also important that they do not overlook important research areas.

Map users outperformed Google users in every parameter: Map users' average score was 84.5%, and they have discovered on average 1.62 seminal papers. Google users achieved a score of 74.2%, and they found 1.2 seminal papers on average. Map users' average recall score was 73.1%, compared to Google's 46.4%.

News Domain. In another study we sought to understand the value of metro maps in helping users understand a complex topic. We took as a test of the deep understanding of a topic the demonstration of an ability to explain that topic to others. We recruited 15 undergraduate students and asked them to write two paragraphs: one summarizing the Haiti earthquake, and one summarizing the Greek debt crisis. For each of the stories, the students were randomly assigned either a metro map or the Google News result page.

We employed crowdworkers on Mechanical Turk to evaluate the paragraphs. At each round, workers were presented with two paragraphs (map user vs. Google News user), and asked to assess which paragraph provided a more complete and coherent picture of the story. 72% of the Greece comparisons preferred map paragraphs, but only 59% of Haiti. After examining the Haiti paragraphs, we found that most map users focused solely on the major storyline (distributing aid). We conclude that maps are more useful for stories without a single dominant storyline.

5. CONCLUSIONS AND FUTURE WORK

We summarized our efforts to develop methods that extract information and construct summarizing metro maps. Given a query, our algorithm generates a concise structured set of storylines which maximizes coverage of salient pieces of information. Most importantly, metro maps explicitly show the relations between lines. We conducted promising pilot user studies in two domains, science and news. The results indicate that our method helps users acquire knowledge efficiently.

In the future, we plan to experiment with richer forms of input, output and interaction models. Possible directions include zooming mechanisms, node-based and line-based feedback, and integration of higher-level semantic features. We believe that this line of work can lead to the fielding of tools that help people navigate and understand ideas, trends, connections and storylines amidst an information explosion.

ACKNOWLEDGMENTS

This work was partially supported by ONR PECASE N000141010672, ARO MURI W911NF0810242, ONR MURI N000141010934 and NSF Career IIS-0644225. Dafna Shahaf was supported in part by Microsoft Research Graduate Fellowship.

REFERENCES

Google Scholar, http://scholar.google.com.

ALLAN, J., GUPTA, R., AND KHANDELWAL, V. 2001. Temporal summaries of new topics. In *SIGIR '01*.

SHAHAF, D., GUESTRIN, C., AND HORVITZ, E. 2012a. Metro maps of science. In *KDD '12*.

SHAHAF, D., GUESTRIN, C., AND HORVITZ, E. 2012b. Trains of thought: Generating information maps. In *WWW '12*.

SWAN, R. AND JENSEN, D. 2000. TimeMines: Constructing Timelines with Statistical Models of Word Usage. In *KDD' 00*.

YAN, R., WAN, X., OTTERBACHER, J., KONG, L., LI, X., AND ZHANG, Y. 2011. Evolutionary timeline summarization: a balanced optimization framework via iterative substitution. In *SIGIR' 11*.

Dafna Shahaf is a postdoctoral fellow at Stanford University. She continues to pursue methods for helping people cut through the explosion of information and find structure in complex topics.

Carlos Guestrin is the Amazon Professor of Machine Learning in the Computer Science and Engineering Department at the University of Washington. His research interests include large-scale machine learning, distributed systems and information retrieval.

Eric Horvitz is a distinguished scientist and co-director of the Microsoft Research lab at Redmond with interests in machine learning, decision making, and human-computer interaction.