Temporal Summarization of Knowledge Evolutionary Trend

Yutao Zhang, Ningping Wang and Jie Tang

Problem

The goal of this work is to produce intuitive summarizations of given evolving research topics. Extracting knowledge concepts, milestones and the backbone of the evolutionary trend. It is essentially a multi-document summarization problem in a dynamic setting. We need to find clusters of knowledge concepts, researchers and publications that are highly indicative of major research topics.

We choose an optimal summarization of the knowledge evolutionary trend base on *content coverage* and *burstiness* of words, the method yields intuitive coarse-level summarization of the temporal dynamics of a given research topic.

Challenges

The temporal summarization task consists a set of unique challenges:

- The research topics are not mutually exclusive. The summarization must consist no only the query it self, but also the related topics.
- To summarize a research topic, we need to find a meaningful description of the topic. However the definition of "meaningfulness" is subjective.
- The arbitrariness of a user given query. Users may want to query unpopular or fine-grained research area which might lead to data sparsity.
- The trade off between conciseness and content coverage of the summarization.

Wisdom of The Crowd

We will leverage user interactions to involve human into the loop. The system will captures the revisions provided by the users and use them to further refine the summarization result.

References

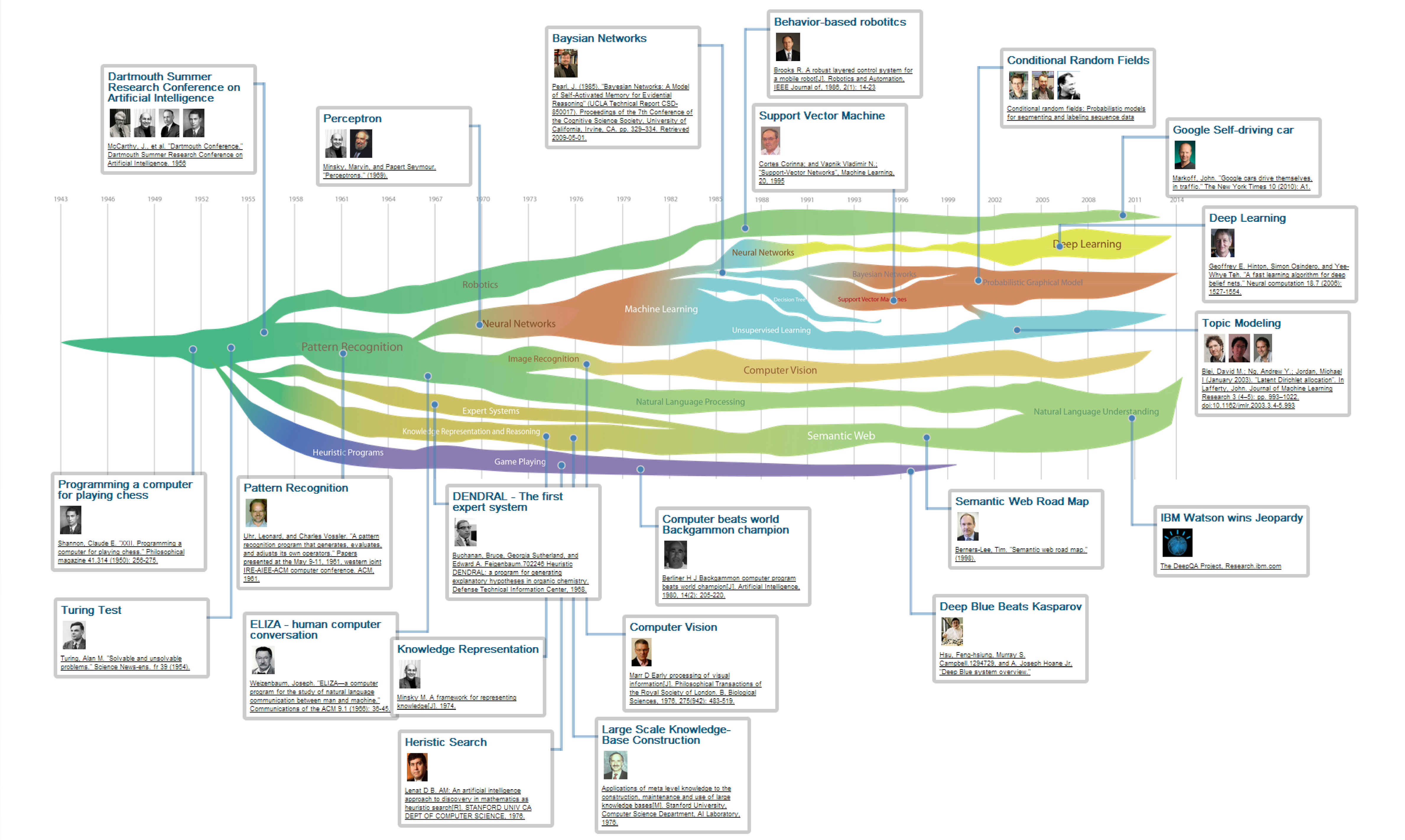
[1] Chi Wang, Xiao Yu, Yanen Li, Chengxiang Zhai, and Jiawei Han Content Coverage Maximization on Word Networks for Hierarchical Topic Summarization. In *CIKM '13*

[2] Chieu, Hai Leong and Lee, Yoong Keok Query Based Event Extraction Along a Timeline In *SIGIR '04*

[3] Cui, Weiwei, Shixia Liu, Li Tan, Conglei Shi, Yangqiu Song, Zekai Gao, Huamin Qu, and Xin Tong. Textflow: Towards better understanding of evolving topics in text. In *Visualization and Computer Graphics, IEEE Transactions on 17, no. 12 (2011): 2412-2421.*

[4] Griffiths, T., M. Jordan, J. Tenenbaum, and David M. Blei. Hierarchical topic models and the nested Chinese restaurant process. In *Advances in neural information processing systems 16 (2004): 106-114.*

Visualization



The evolutionary trend of Artificial Intelligence.

Method

We introduce a community-based summarization as we first find a core community (a group of experts) around the topic, and then aggregate each member's research work as the research area. To avoid some authoritative experts dominates the research area and introduce bias, we normalize each member's contribution averagely.

With the retrieved document collection, we extract the knowledge concepts mentioned in each documents and construct a dynamic mutual information graph. The following task is to choose a set of high level concepts best summarize the whole research area. We use two measures: *Coverage* and *Burstiness* to choose the concepts of interest.

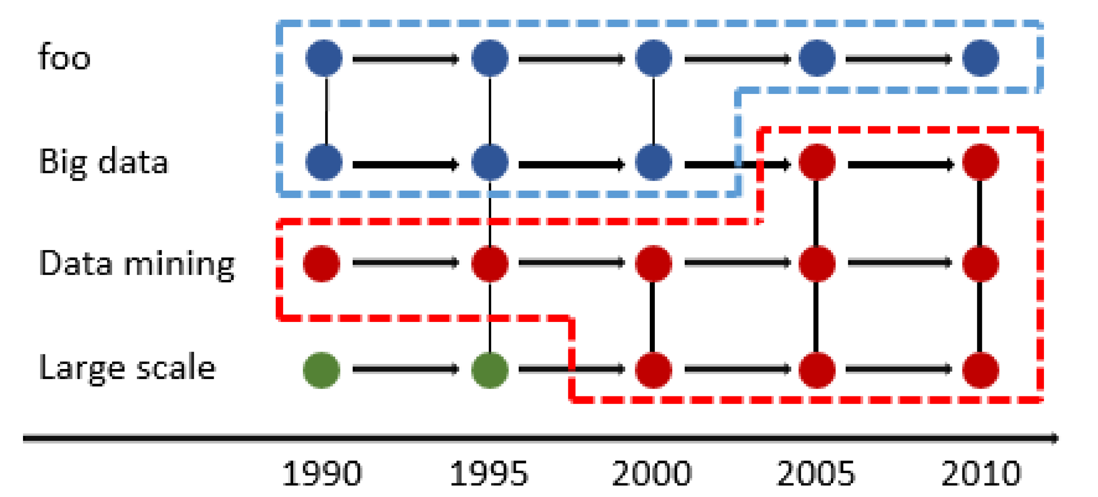
We use a influence maximization based model to model information coverage of knowledge concepts. Formally we choose a set of k concepts that maximize the content coverage. Meanwhile, we model burstiness by assuming the arrival of words as a unknown binomial distribution, and use χ^2 tests to check for significant association between word and time periods. We calculate the contingency table as following and $\chi^2 = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)}$.

-	W	$\bar{W}$
t	a	b
$< t$	c	d

After choosing a set of summarization words, we further grouping the rest of the words into clusters, and finally we can use the clusters over time to indict evolving trend and generates the highly intuitive visualization.

Representation

We represent the temporal dynamics of knowledge concepts in a dynamic mutual information graph. For each time slide, we create a node for each related knowledge concepts.



There are two types of edges in the graph:

- The concepts in the same time slide are connected with mutual info edge. For an arbitrary edge (u, v), the influence probability from u to v is $P(u, v) = \frac{I(u,v)}{I(u,u)}$, where $I(u, v)$ indicates the mutual information between u and v.
- Nodes corresponding to the same knowledge concept within two adjacent time slides are connected by an evolving edge, indicates the knowledge concept is evolved from itself in the last time slide. The weight of the evolving edge is a parameter controlling the temporal smoothness of the summarization.

Online Demo

The online demo is available at `http://arnetminer.org/event/aihistory`