Temporal Summarization of Knowledge Evolutionary Trend

Yutao Zhang, Ningping Wang and Jie Tang

Problem

The goal of this work is to find a optimal summarization of the evolving research field. Extracting knowledge concepts, milestones and the backbone of the evolutionary trend. It is essentially a heterogeneous graph 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 summarize the knowledge evolutionary trend base on MDL principle, the method yields intuitive coarse-level summarization of the temporal dynamics of a given research topic.

Challenges

- For an arbitrary research topic given by the user. Thus, we introduce the community-based summarization as we first find the core community around this topic, and aggregate each member's research
- Query based system.
- Core-Community.
- important is subject
- **Data sparsity** As a query based system, we need to summarize arbitrary research topics given by the user.

MDL Principle

The solution is constrained by a shape model

$$M(\mathbf{\Theta}) = (m_1(\mathbf{\Theta}), \ldots, m_N(\mathbf{\Theta})) \qquad (1)$$
$$m_i : \mathbb{R}^{N_\Theta} \to \mathbb{R}^2$$

mapping model parameters $\mathbf{\Theta}$ to image positions $m_i(\mathbf{\Theta})$. For each fiducial point m_i a set of candidate positions

$$L_i = \{l_i^1, l_i^2, \ldots\} \qquad l_i^j \in \mathbb{R}^2 \qquad (2)$$

is detected in the image. The task is to assign to every model vertex one of the candidate positions such that the shape model can be best fit to the selection $\mathbf{S}$, written as a tuple

$$\mathbf{S} = (j_1, j_2, \ldots, j_N) \qquad j_i \in \mathbb{N}, \qquad (3)$$

where j_i is the index of a candidate of landmark i.

So we minimize the distance between the shape model and the image landmarks:

$$\mathbf{S} = \arg\min_{\mathbf{S}=(j_1\ldots j_N)} f(\mathbf{S})$$

$$f(\mathbf{S}) = \min_{\Theta} \sum_i \rho\left(\left\| m_i(\Theta) - l_i^{j_i} \right\|\right) \qquad . \qquad (4)$$

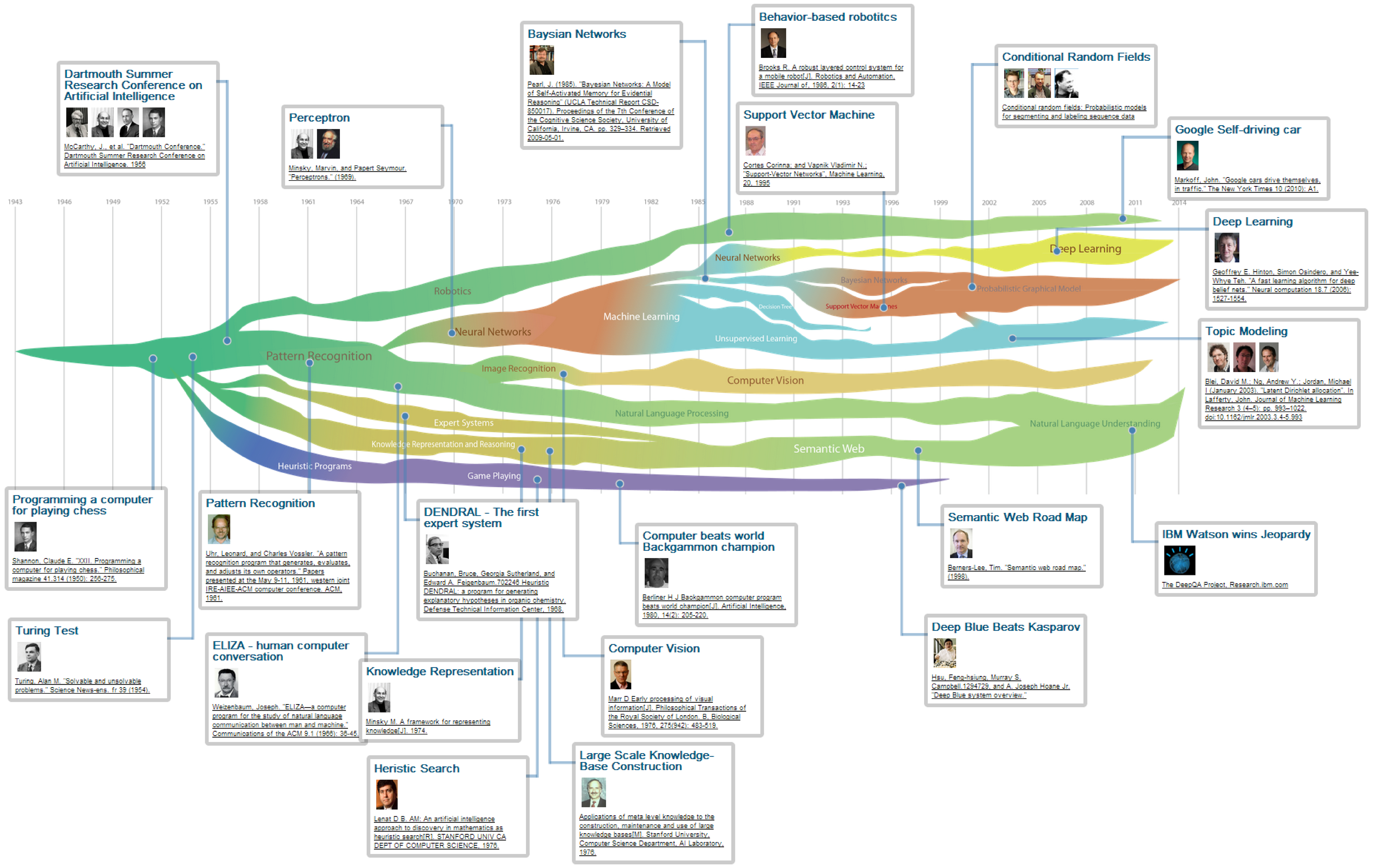
Where $\rho : \mathbb{R} \to \mathbb{R}$ is a robust function, allowing us to handle missing detections, and points which are invisible due to occlusion.

References

[1] B. Amberg, T. Vetter. Optimal Landmark Detection using Shape Models and Branch and Bound. In ICCV '11

[1] B. Amberg, T. Vetter. Optimal Landmark Detection using Shape Models and Branch and Bound. In ICCV '11

Visualization



The summarization of the evolutionary trend of Artificial Intelligence.

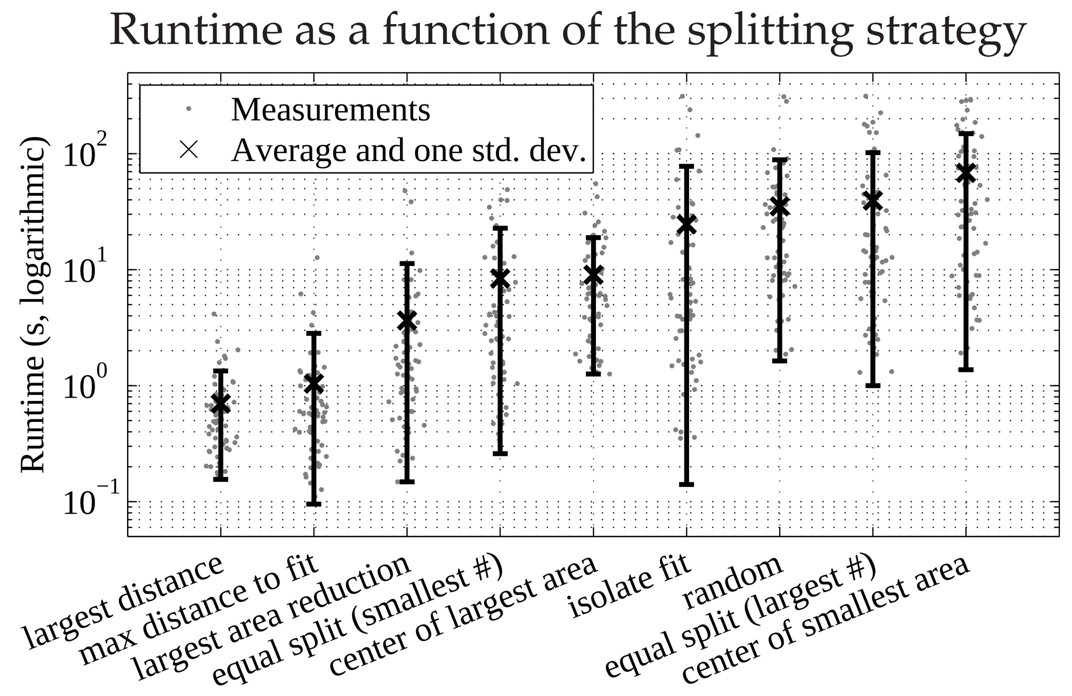
Method

This discrete optimization is solved by Branch and Bound, which is a method to minimize a function over a set. It requires us to (1) efficiently specify solution subsets, (2) determine a lower bound on the minimal cost of the solutions within a subset, and (3) specify a strategy to split a solution subset into two new subsets.

The ingredients in our case are:

1. Solution subsets are created by taking subsets of landmark candidates, and considering the Kartesian product of all selected landmark candidates
2. We bound the cost for such a solution set by taking for each landmark the minimal distance to the convex hull of the selected candidates
3. We found that splitting landmark candidates such that the convex hull of the resulting two landmark candidates are as distant as possible is most effective.

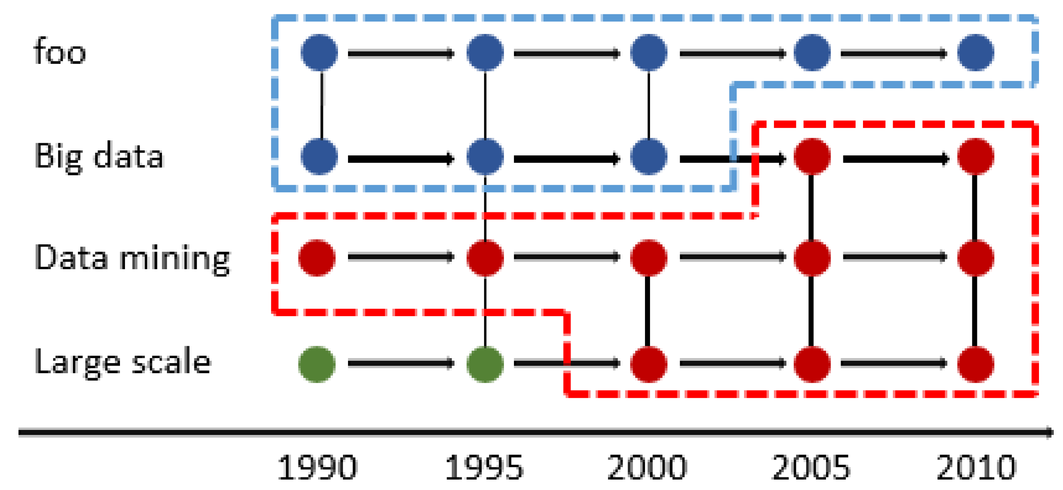
Splitting Strategy



Different splitting strategies result in vastly different performance. Note that 'split into equal sized problems' is one of the worst strategies for branch and bound.

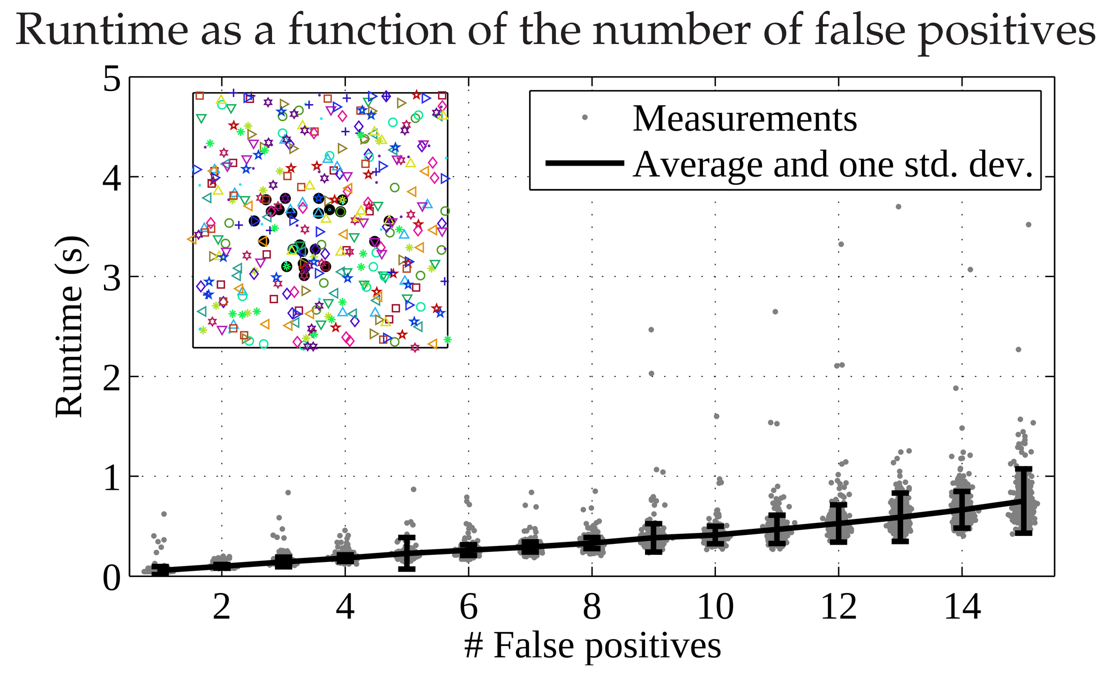
Representation

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 same knowledge concept within two adjacent time slides is 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.

Wisdom of the Crowd



Online Demo

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