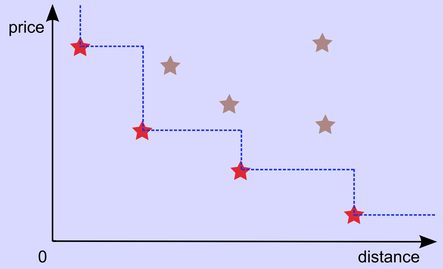
Interactive Ranking of Skylines using Machine Learning Techniques

Weiwei Cheng, Eyke Hüllermeier, Bernhard Seeger, Ilya Vladimirskiy
Department of Mathematics and Computer Science
Marburg University, Germany

Ranking the Skyline

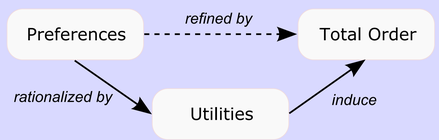


The *skyline* operator maps a finite set $\mathcal{O}$ of objects, each characterized in terms of a fixed number of features (criteria), to the subset of Pareto-optimal elements:

$$P(\mathcal{O}) \stackrel{\text{df}}{=} \left\{ \mathbf{o} \in \mathcal{O} \,|\, \{\mathbf{o}' \in \mathcal{O} \,|\, \mathbf{o} \prec \mathbf{o}'\} = \emptyset \right\}$$

Important problem: $P(\mathcal{O})$ may become huge, especially in high dimensions!

Ranking the skyline via a (latent) *utility function*:



- A utility function $U(\cdot)$ assigns a real utility degree to each object $\mathbf{a} = (a_1 \ldots a_d) \in \mathcal{O}$; $U(\mathbf{a}) < U(\mathbf{b})$ means that the user strictly prefers $\mathbf{b}$ to $\mathbf{a}$.
- Utility degrees induce a total order; thus, a ranking can be presented instead of an unsorted answer set.
- User feedback is used to improve ranking quality.

Algorithm Design

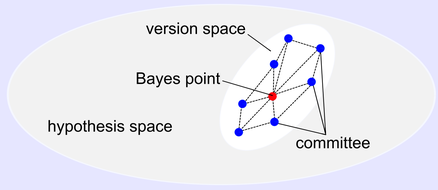
Base Learner: Noise tolerant perceptron with margin.

Training Data: A set of revealed (pairwise) preferences $\mathbf{a} \prec \mathbf{b}$, turned into positive and negative examples for classification.

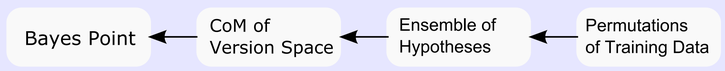
Monotonicity: $\mathbf{a} \geq \mathbf{b} \Rightarrow U(\mathbf{a}) \geq U(\mathbf{b})$ must be guaranteed for all $\mathbf{a}, \mathbf{b} \in \mathcal{O}$.

Utility: Linear model $U(\mathbf{a}) = \langle \mathbf{w}, \mathbf{a} \rangle = w_1 a_1 + \ldots + w_d a_d$ (monotonicity holds if $\mathbf{w} \geq 0$) and kernalized version.

Bayes Point Machine:



Approximation of the Bayes point by the center of mass of version space.

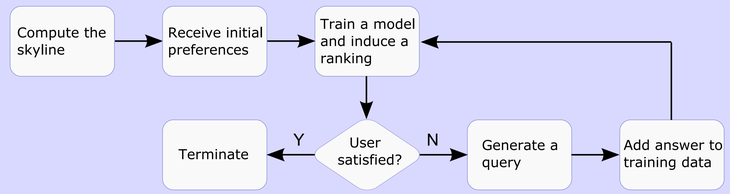


Active Learning Strategy:

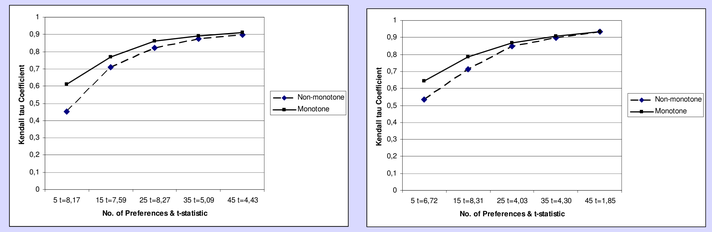
1. Constitute a committee of learners.
2. Find two maximally conflicting learners.
3. For each learner, generate a corresponding ranking. Return the first discordant pair as a query. Add the answer to the preference set.
4. Retrain the committee on the enlarged preference set and go to step 2.

Experimental Results

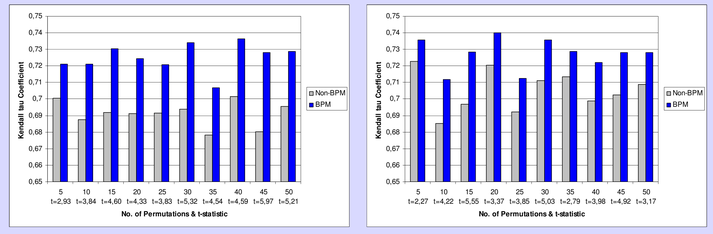
Workflow:



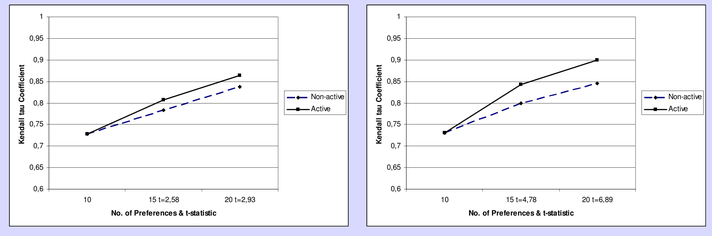
Monotone vs. non-monotone learning:



Ensemble (Bayes point machine) vs. single learner:



Active vs. non-active learning:



Poster presented at KDML 2007, *Workshop Knowledge Discovery, Data Mining, and Machine Learning*, Halle/Saale, September 2007.