

Tuesday: Foundations

Learning objectives of the day:

- 1) Become familiar with the theoretical and algorithmic foundations of three selected topics.
- 2) See how proper modelling and engineering can be applied.

Keywords: algorithms, spatial data structures, clustering, probabilistic graphical models, Bayesian methods, trajectory data.

I. PRIORITIZED GEOGRAPHIC SEARCH

The first lecture will be given by Stefan Funke and Sabine Storandt. With the planet-wide availability of geo-referenced data there is a need for efficient retrieval methods. But queries as e.g. 'European cities with a historic center' result in too many hits to be visualized or even just listed in a short amount of time. Also the user most likely does not want to be shown a cluttered map with all the valid results but rather prefers to see the most important hits, where importance could e.g. be defined as the size of the respective cities. So given a (huge) set of prioritized points of interest, we want to answer range queries equipped with an additional minimum priority (e.g. more than 200,000 inhabitants). Ideally, when answering such a query as few points as possible outside the desired range and with a priority below the minimum property should be touched in order to enable a fast response.

In this lecture we will discuss data structures for that purpose. In the accompanying exercise session, attendees will implement one of those data structures and evaluate them using real-world data.

II. SUBSPACE CLUSTERING AND PROBABILISTIC GRAPHICAL MODELS FOR DATA PROCESSING

The second lecture, by Hanno Ackermann and Michael Yang, will focus on subspace clustering methods and probabilistic graphical models (PGMs) for data processing, and in particular on classifying and clustering of spatiotemporal data. Many real world problems in computer vision require to reason about highly uncertain, structured data, and draw global insight from local observations. PGMs allow addressing these challenges in a unified framework. Several applications using Markov random fields will be presented.

III. HYPTRAILS

In the third lecture, Florian Lemmerich will present HypTrails, a general approach for comparing a set of hypotheses about human trails on the Web, where hypotheses represent beliefs about transitions between states. Our approach utilizes Markov chain models with Bayesian inference. The main idea is to incorporate hypotheses as informative Dirichlet priors and to leverage the sensitivity of Bayes factors on the prior for comparing hypotheses with each other. For eliciting Dirichlet priors from hypotheses, we present an adaption of the so-called (trial) roulette method. We demonstrate the general mechanics and applicability of HypTrails

IV. EXERCISES

Exercises on these topics will be held Tuesday afternoon, allowing participants to get an in-depth view into their topic of choice.