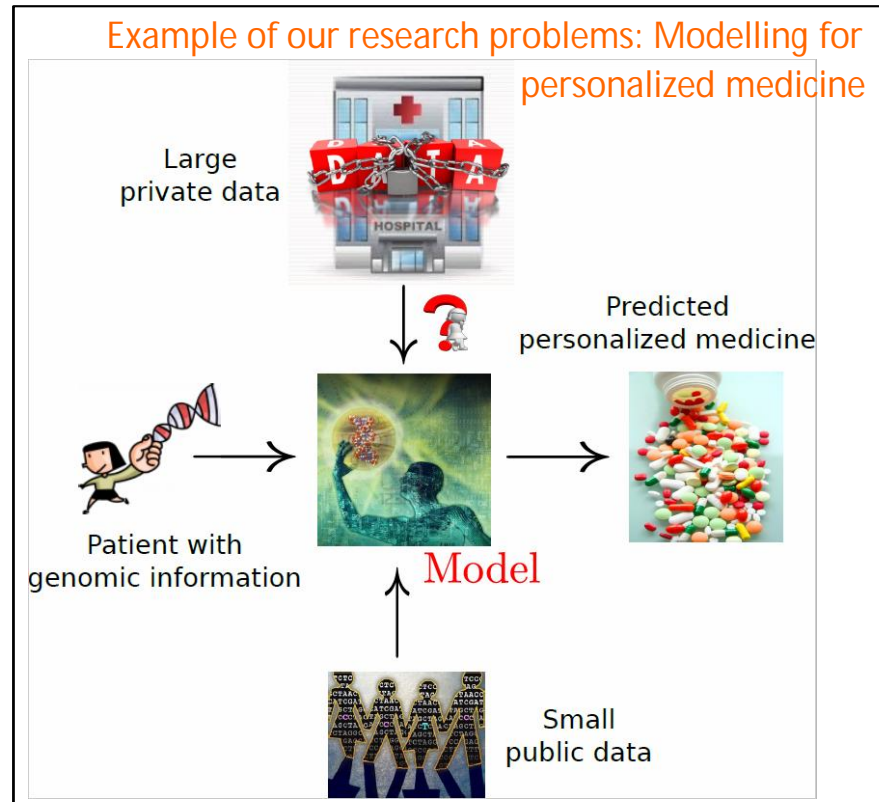


Probabilistic Machine Learning

We develop new methods for probabilistic modeling, Bayesian inference and machine learning.

Keywords:

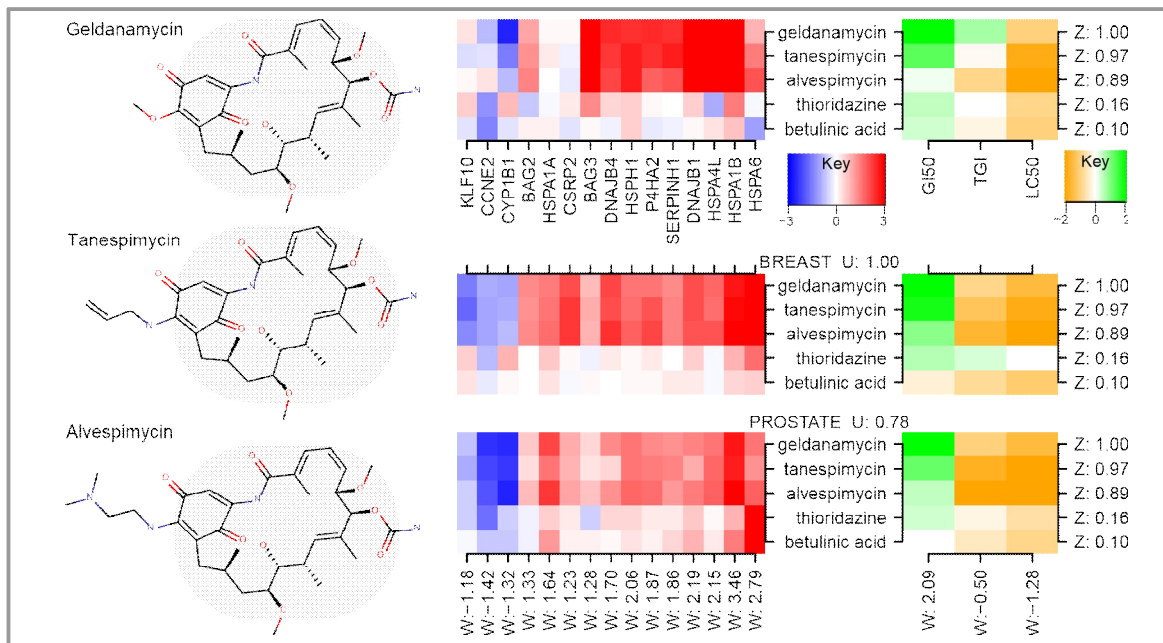
- learning from multiple data sources
- Gaussian processes
- data visualization
- retrieval of relevant data
- machine learning for user interaction
- personalized medicine
- brain signal analysis and neuroinformatics



Modeling Multiple Data Sources

We develop models for inferring relations of multiple data sources, allowing detection of shared effects and robust predictions of missing data in any source.

Example: drug toxicity prediction



1. Gather existing drug toxicity measurements on a variety of drugs and cancer cell lines

2. Pair the toxicity data with gene expression data and structural drug descriptors

3. Learn relations between the data sources; such as heat-shock protein response (left)

4. Use the relations affecting drug toxicity to predict missing drug-cancer pairs

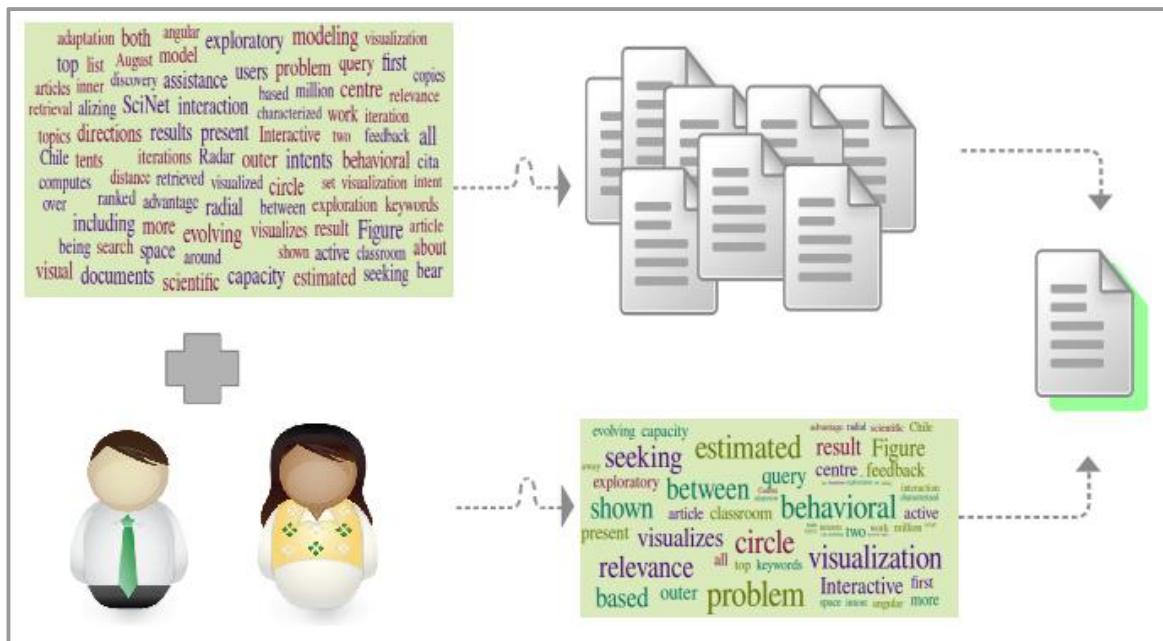
Tools and applications:

- generative models
- drug response studies
- non-linear regression
- genome-wide association studies

Prior Knowledge Elicitation

We study prediction methods that use interactive elicitation of prior user knowledge. Wide range of applications (personalized medicine, recommendation, etc.).

Example: interactive intent modeling and article recommendation



1. Start from a collection keywords and their relevance for the user.
2. Preliminary prediction of a set of potentially relevant articles.
3. Elicit the prior knowledge of expert users: which keywords are more relevant?
4. Adjust the keyword relevance according to the feedback from the expert users.
5. Re-learn the prediction model and recommend the best article according to the updated user intent estimation.

Machine Learning tools:

- probabilistic modeling
- sequential decision making
- exploration/exploitation
- adaptive interactive system

Inference and Data Integration

We develop methods for inference on complex and intractable models and integration of multiple datasets.

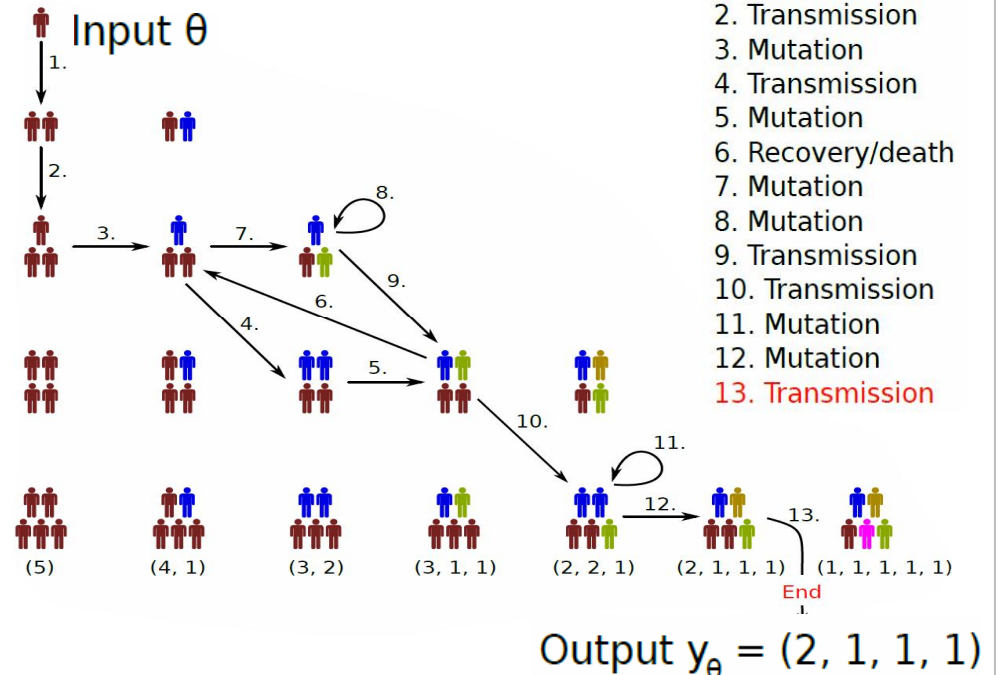
Research topics include

- Bayesian probability theory and methodology
- Model selection and assessment
- Distributed Bayesian computation
- Likelihood-free inference
- Privacy aware learning
- Phase transition behavior in statistical inference

Key application domains:

- Bioinformatics
- Epidemiology
- Personalized medicine
- Brain signal analysis
- Information retrieval

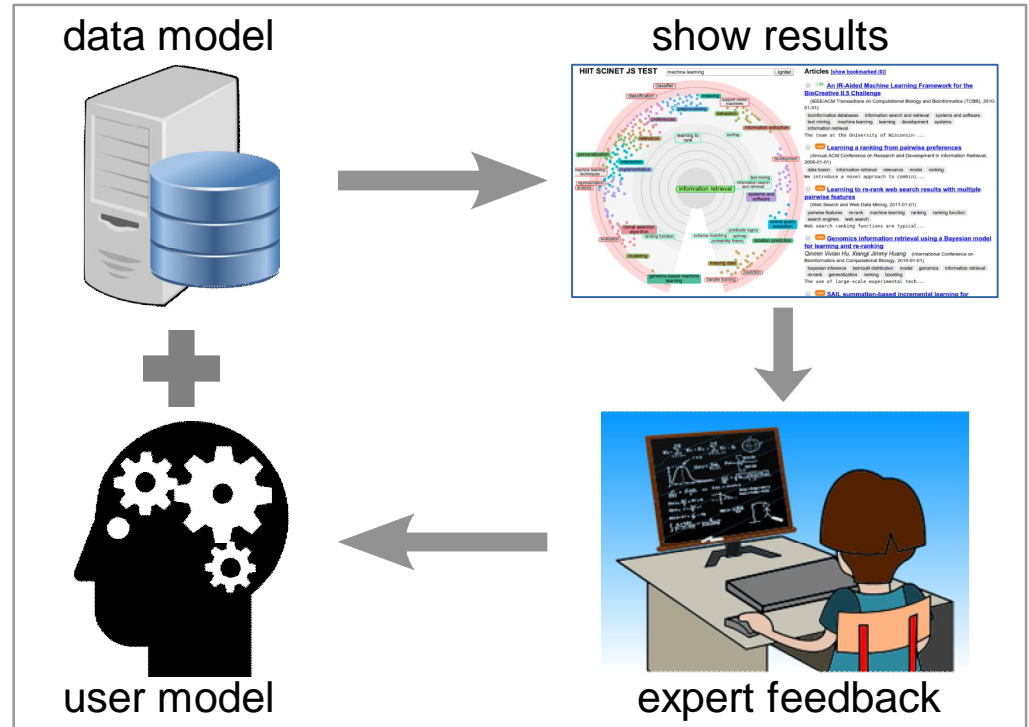
Example: Likelihood-free inference of tuberculosis transmission parameters



Interactive Information Visualization

We visualise data, according to a model of user's knowledge and intentions, learned through user interaction and feedback.

1. The user starts new search or gives feedback to previous results.
2. Based on user's interactions learn the model of user's intentions.
3. Update the data model according to the user's needs.
4. Based on user and data models create a human readable representation of the data.



Applications:

- personalized cancer medicine
- search in private data
- search for scientific documents