

Advanced Data Mining Methods

Data Mining:
Data Mining Methods
with Dr. Qin Lv



Master of Science in Data Science
UNIVERSITY OF COLORADO BOULDER



Learning objective: Identify some advanced data mining methods and research frontiers of data mining.

Advanced Data Mining

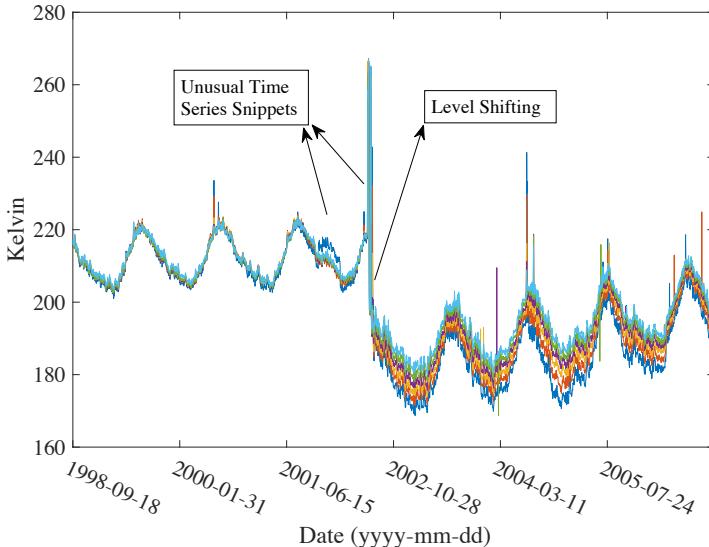
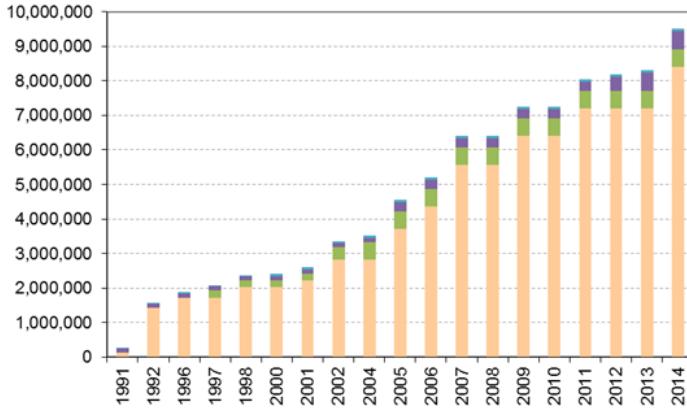
- Sequence data, time series
- Graph data, social networks
- Web data, data fusion
- Research frontiers

Sequence Data

- Ordered list, with or without time
 - E.g., biological sequence, stock price, trajectory
- Sequential pattern mining
 - Frequent subsequence
 - Pairwise sequence alignment: e.g., BLAST
 - Sequence modeling/prediction: e.g., Markov chain model

Time Series Data

- Changes over time
- Combined signal of
 - T: overall trend
 - C: cyclic patterns
 - R: random noise
 - A: Anomalies
 - ...



Graph Data

- $G = (V, E)$: relationships between entities
 - E.g., movies, road network, (un)directed & weighted links
- Graph mining
 - Frequent subgraph
 - Anomaly detection
 - Graph modeling
 - Link prediction



Online Social Network (OSN) Data

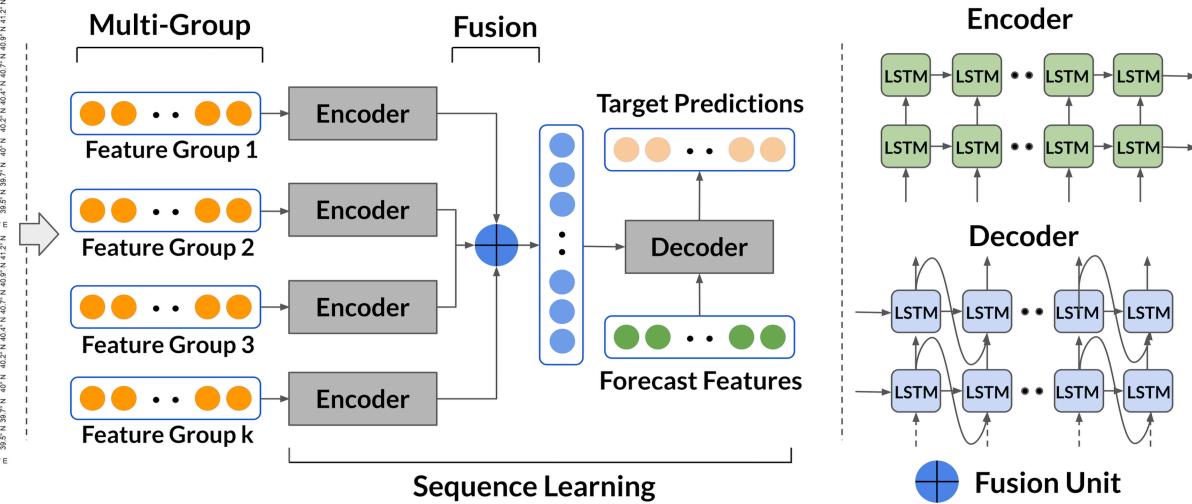
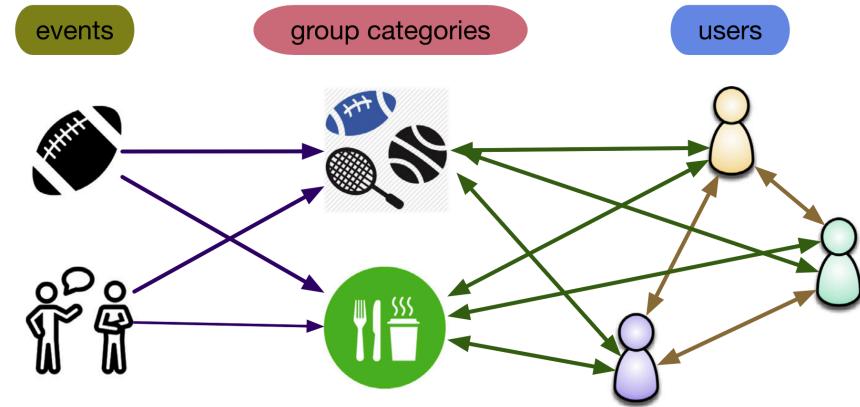
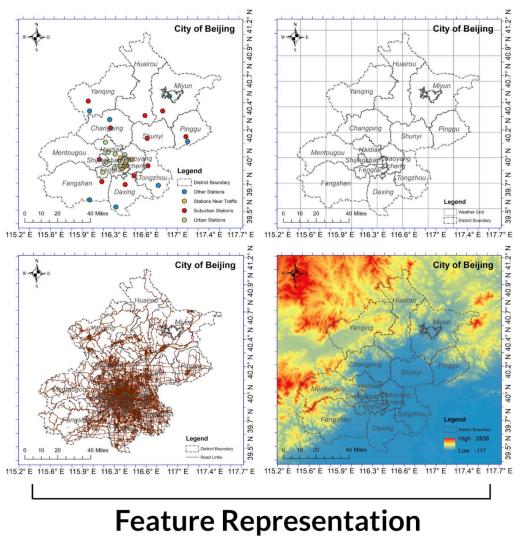
- Users/groups, content, interactions
- OSN mining
 - OSN modeling, community detection
 - Topic modeling, sentiment analysis
 - Information diffusion, recommendation
 - Cybersafety: malicious accounts & behaviors

Web Data

- Text, image, audio, video, hyperlinks
 - E.g., image clustering: image content, text description, hyperlinked webpages
- Search engines
- Recommender systems
- Knowledge graphs

Data Fusion

➤ Multi-modal data



Research Frontiers

- Active research in the field
 - Both in academia and real-world application domains
- ACM Special Interest Group in Knowledge Discovery & Data Mining
 - <https://www.kdd.org>
 - Community for data mining, data science and analytics

Annual KDD Conference

- Keynote talks
- Research track
- Applied data science track
- Tutorials: lecture, hands-on
- Workshops
- KDD Cup

Types of Data

- Text, Web, and social media data
- Time series and data streams
- Graph data
- Images, audio, video, geospatial data
- Multi-modal data

Data Mining Methods

- Unsupervised learning
- Semi-supervised learning
- Transfer learning, multitask learning
- Deep learning, reinforcement learning
- Recommender systems

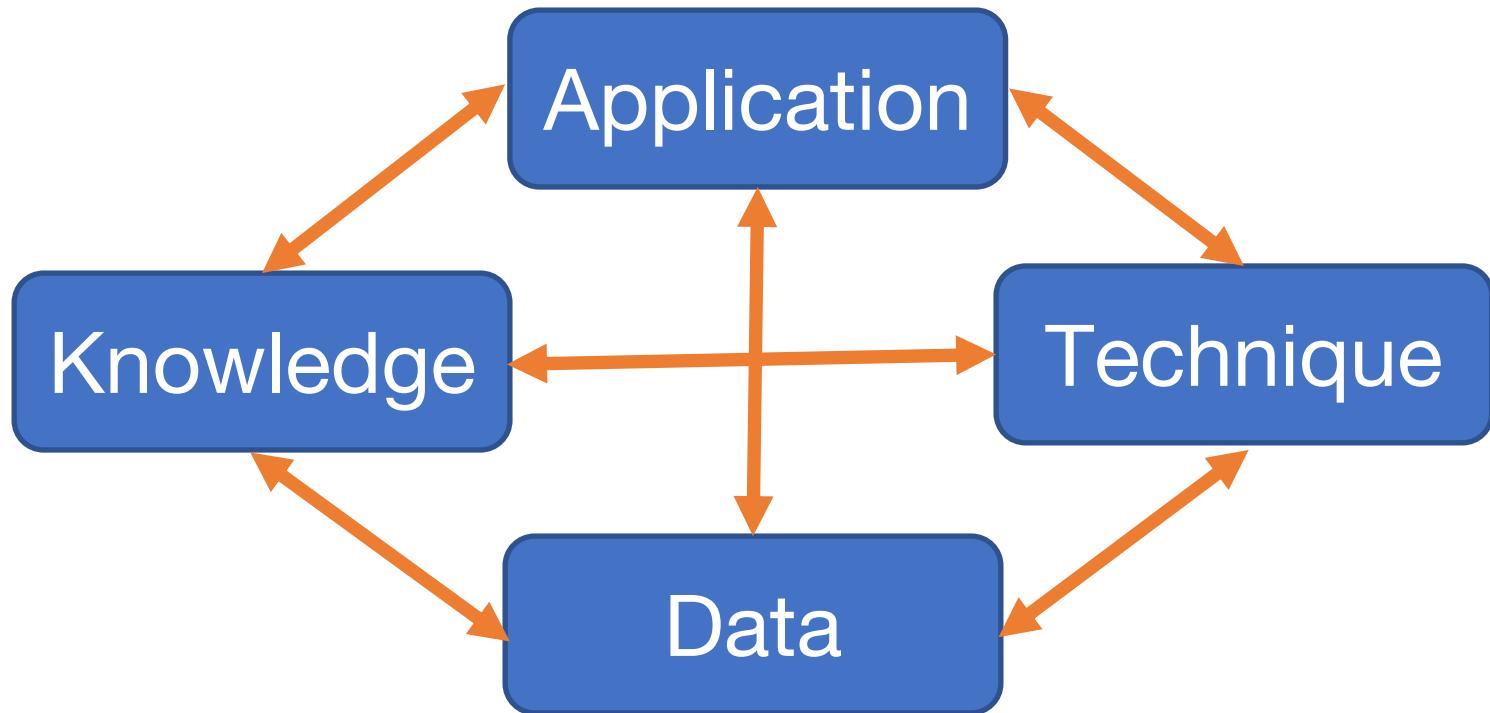
Application Domains

- Healthcare
- Business intelligence
- Earth, environment
- Scientific discovery
- Industry AI

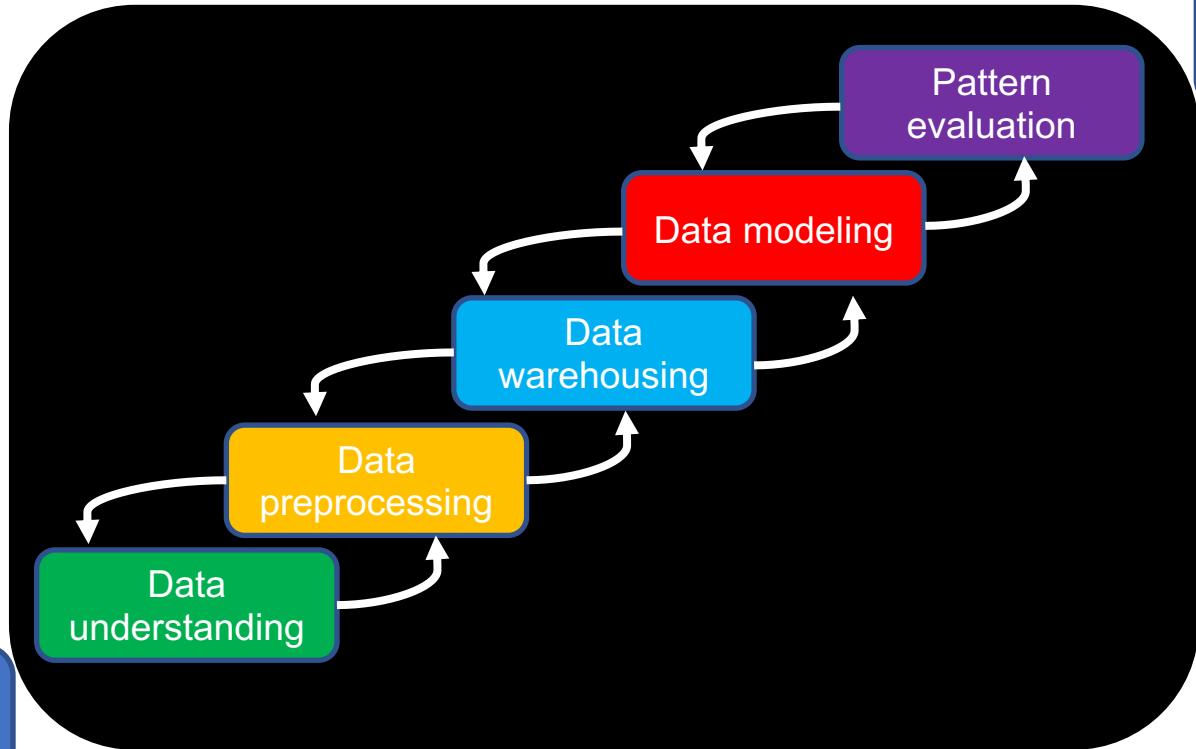
Beyond Accuracy

- Efficiency, scalability
- Interpretability
- Automation
- Fairness, equity, justice
- Social impact, social good

Data Mining: Four Views



Data Mining Pipeline



Application

Knowledge

Technique

Data

Data Mining Methods

- Frequent pattern analysis
- Classification
- Clustering
- Outlier analysis
- Advanced methods & research frontiers