CS412
Summary and Beyond

Kevin C. Chang

Knowledge Discovery (KDD) Process

Knowledge

Pattern Evaluation

Data Mining

Task-relevant Data

Selection

Data Warehouse

Data Cleaning

Data Integration

Databases

2
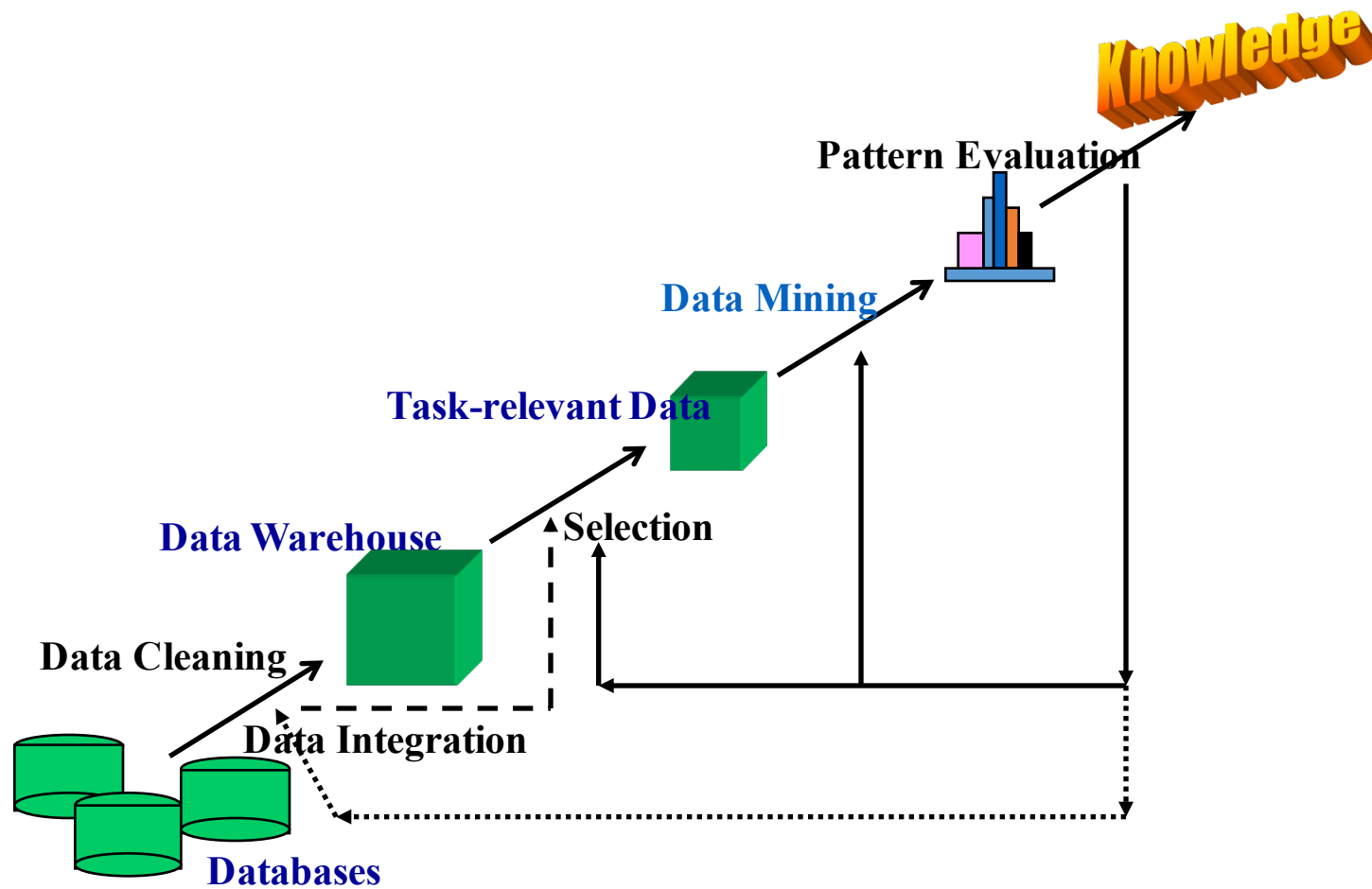
What have we studied?

- Chapter 2: Know Your Data
- Chapter 3: Data Preprocessing
- Chapter 4: Data Warehousing & OLAP
- Chapter 5: Data Cube Technology
- Chapter 6: Mining Frequent Patterns and Associations: Basic Concepts
- Chapter 7: Mining Frequent Patterns, Associations: Advanced Methods
- Chapter 8: Classification: Basic Concepts
- Chapter 9: Classification: Advanced Methods
- Chapter 10: Cluster Analysis: Basic Concepts

Chapter 2: Getting to Know Your Data

- Data Objects and Attribute Types

- Basic Statistical Descriptions of Data

- Data Visualization

- Measuring Data Similarity and Dissimilarity

- Summary

Chapter 3: Data Preprocessing

- Data Preprocessing: An Overview

  - Data Quality

  - Major Tasks in Data Preprocessing

- Data Cleaning

- Data Integration

- Data Reduction

- Data Transformation and Data Discretization

- Summary

Chapter 4: Data Warehousing and On-line Analytical Processing

- Data Warehouse: Basic Concepts

- Data Warehouse Modeling: Data Cube and OLAP

- Data Warehouse Design and Usage

- Data Warehouse Implementation

- Summary

Chapter 5: Data Cube Technology

- Data Cube Computation: Preliminary Concepts

- Data Cube Computation Methods

- Processing Advanced Queries by Exploring Data Cube Technology

- Multidimensional Data Analysis in Cube Space

- Summary

Chapter 6: Mining Frequent Patterns, Association and Correlations: Basic Concepts and Methods

- Basic Concepts

- Frequent Itemset Mining Methods

- Which Patterns Are Interesting?—Pattern Evaluation Methods

- Summary

Chapter 7 : Advanced Frequent Pattern Mining

- Pattern Mining: A Road Map

- ~~Pattern Mining in Multi-Level, Multi-Dimensional Space~~

- Constraint-Based Frequent Pattern Mining

- ~~Mining High-Dimensional Data and Colossal Patterns~~

- ~~Mining Compressed or Approximate Patterns~~

- Sequential Pattern Mining

- Graph Pattern Mining

- Summary

Chapter 8. Classification: Basic Concepts

- Classification: Basic Concepts

- Decision Tree Induction

- Bayes Classification Methods

- ~~Rule-Based Classification~~

- Model Evaluation and Selection

- Techniques to Improve Classification Accuracy: Ensemble Methods

- Summary

Chapter 9. Classification: Advanced Methods

- ~~Bayesian Belief Networks~~

- ~~Classification by Backpropagation~~

- Perceptron, Support Vector Machines

- ~~Classification by Using Frequent Patterns~~

- Lazy Learners (or Learning from Your Neighbors)

- ~~Other Classification Methods~~

- Additional Topics Regarding Classification

- Summary

Chapter 10. Cluster Analysis: Basic Concepts and Methods

- Cluster Analysis: Basic Concepts

- Partitioning Methods

- Hierarchical Methods

- Density-Based Methods

- Grid-Based Methods

- Evaluation of Clustering

- Summary

Where to go from here?

Where to go from here?

- Study more

- Do something

Study More

- CS411: Database Systems
- CS511: Advanced Data Management
- CS410: Intro to Text Information Systems
- CS510: Advanced Topics in Information Retrieval
- CS412: Introduction to Data Mining
- CS512: Data Mining: Principles and Algorithms
- CS446: Machine Learning
- CS447: Natural Language Processing
- CS546: Machine Learning in NLP

Do Something

- Build a project

- Do research

- Find a job

Happy Holidays