

Technology
Science
Information
Networks
Computing



Lecturer: Ting Wang (王挺)

利物浦大学计算机博士

清华大学计算机博士后

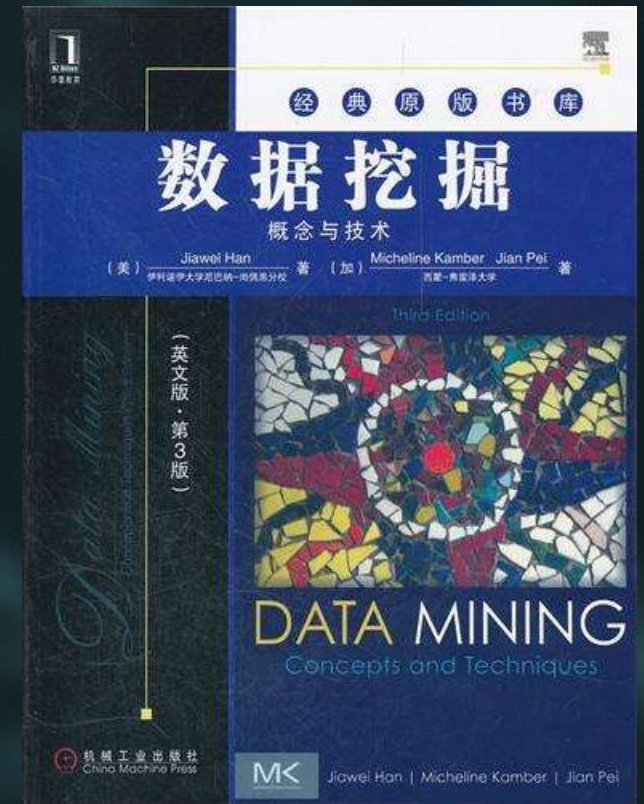
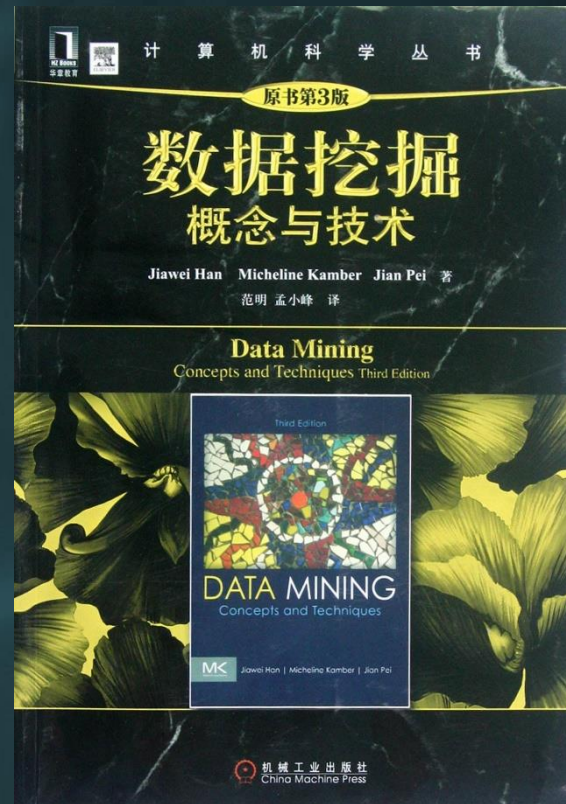
电子信息技术高级工程师

上海外国语大学网络与新媒体副教授

浙江清华长三角研究院海纳认知与智能研究中心主任

Chapter 7

Advanced Pattern Mining



Chapter 7: Advanced Pattern Mining

1. rare pattern (infrequent pattern)
occur rarely but are of special interest

a pattern with a frequency support that is below (or far below) a user-specified minimum support threshold.

Example:

Sales of diamond watches



Approaches:

- mining **multilevel** Associations(7.2.1)
- constraint-based pattern mining

Chapter 7: Advanced Pattern Mining

2. negative pattern

Patterns with components that exhibit negatively correlated behavior

浙C 88888的故事：有意义的负模式

Definition 7.1: If itemsets X and Y are both frequent but rarely occur together (i.e., $sup(X \cup Y) < sup(X) \times sup(Y)$), then itemsets X and Y are **negatively correlated**, and the pattern $X \cup Y$ is a **negatively correlated pattern**. If $sup(X \cup Y) \ll sup(X) \times sup(Y)$, then X and Y are **strongly negatively correlated**, and the pattern $X \cup Y$ is a **strongly negatively correlated pattern**. □

✘ Null-invariant
零不变性

Definition 7.2: If X and Y are strongly negatively correlated, then

$$sup(X \cup \bar{Y}) \times sup(\bar{X} \cup Y) \gg sup(X \cup Y) \times sup(\bar{X} \cup \bar{Y}).$$

Is this measure null-invariant? □

✘ Null-invariant
零不变性

Definition 7.3: Suppose that itemsets X and Y are both frequent, that is, $sup(X) \geq min_sup$ and $sup(Y) \geq min_sup$, where min_sup is the minimum support threshold. If $(P(X|Y) + P(Y|X))/2 < \epsilon$, where ϵ is a negative pattern threshold, then pattern $X \cup Y$ is a **negatively correlated pattern**. □



Chapter 7: Advanced Pattern Mining

3. Constraint-based frequent pattern

mining process toward patterns that match users' intuition or satisfy certain constraints

Types:

- Knowledge type constraint:
 - classification, association, etc.
- Data constraint — using SQL-like queries
 - find product pairs sold together in stores in Chicago this year
- Dimension/level constraint
 - in relevance to region, price, brand, customer category
- Rule (or pattern) constraint
 - small sales (price < \$10) triggers big sales (sum > \$200)
- Interestingness constraint
 - strong rules: $\text{min_support} \geq 3\%$, $\text{min_confidence} \geq 60\%$



Next >> Chapter 8

www.wangting.ac.cn