AI

Contents:

- ☐ 10.1. Classification
- ☐ 10.2. Regression
- ☐ 10.3. Clustering
- ☐ 10.4. Ranking
- ☐ 10.5. Dimensionality Reduction

Ranking

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

What is Ranking 什么是排名

☐ A longer description 较长描述
A ranking is a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second.
  排名是一组项之间的关系，即对于任意两个项，满足第一个"排名高于"、"排名低于"或"排名等于"第二个。

☐ A shorter description 较短描述
The data transformation in which numerical or ordinal values are replaced by their rank.
  排名是一种数据转换，其中数值或者顺序值由其排名来代替。

☐ A very short description 极简描述
To order items according to some criterion.
  依据某种准则整理数据项。

Contents:

A Formal Description of Ranking 一种排名的形式化描述

Let $\mathcal{X}$ denote input space, D an unknown distribution over $\mathcal{X} \times \mathcal{X}$.

设$\mathcal{X}$表示输入空间，D是$\mathcal{X} \times \mathcal{X}$ 上的未知分布。

☐ Target ranking function: 目标排名函数：

$$f : \mathcal{X} \times \mathcal{X} \to \mathcal{Y} = \{-1, 0, +1\}$$

where   其中
- $f(x, x') = +1$, if x is ranked higher than x',   若x排名高于x'，
- $f(x, x') = -1$, if x is ranked lower than x',   若x排名低于x'，
- $f(x, x') = 0$, if both x and x' has same ranking.   若x与x'二者排名相同。

☐ Training data: 训练数据

$$\mathcal{S} = \{(x^{(i)}, x'^{(i)}, y^{(j)}) \mid y^{(j)} = f(x^{(i)}, x'^{(i)}) \in \mathcal{Y}, i \in [1, m] , j \in [1, 3]\}$$

A Formal Description of Ranking 一种排名的形式化描述

□ Ranking problem: 排名问题
Given a hypothesis set H of functions mapping $\mathcal{X} \times \mathcal{X}$ to $\mathcal{Y} = \{-1, 0, +1\}$, to select a hypothesis $h \in H$ with the target function f:

给定一个将$\mathcal{X} \times \mathcal{X}$映射到$\mathcal{Y} = \{-1, 0, +1\}$的假设函数集$H$，选择一个具有目标函数$f$的假设$h \in H$：

■ small expected generalization error: 最小预期泛化错误：

$$R(h) = \Pr_{(x, x')}[f(x, x') \neq 0 \wedge (f(x, x')(h(x') - h(x)) \leq 0)]$$

■ empirical pairwise misranking error: 经验性成对误排名错误：

$$\hat{R}(h) = \frac{1}{m} \sum_{i=1}^{m} 1\left((y^{(i)} \neq 0) \wedge (y^{(i)}(h(x'^{(i)}) - h(x^{(i)})) \leq 0)\right)$$

AI

Contents:

- ☐ 10.4.1. How Ranking Works
- ☐ 10.4.2. Major Approaches of Ranking
- ☐ 10.4.3. Applications and Algorithms

Typical Approaches of Ranking 典型的排名方法

☐ 1) Score-based approach 基于分值方法

■ The predictor is a real-valued function, called *scoring function*.
该预测器是一个实数函数，称为分值函数。

■ The scores assigned to input points by this function determine their ranking.
由该函数分派给输入数据点的分值决定其排名。

■ This approach is the most widely explored one.
这种方法是研究得最多的一种。

☐ 2) Preference-based approach 基于偏好方法

■ The predictor is a *preference function*.
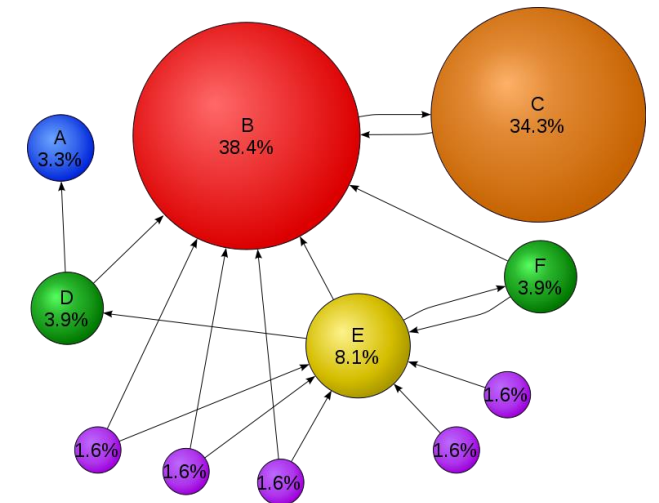该预测器是一个偏好函数。

Contents:

☐ 10.4.1. How Ranking Works

☐ 10.4.2. Major Approaches of Ranking

☐ 10.4.3. Applications and Algorithms

Typical Applications of Ranking 排名的典型应用

☐ In information retrieval   信息检索领域
- Search engine   搜索引擎
- Document retrieval   文档检索
- Collaborative filtering   协同式过滤
- Sentiment analysis   情感分析
- Computational advertising   计算广告学

☐ In other areas   其它领域
- Machine translation   机器翻译
- Recommender systems   推荐系统
- Computational biology   计算生物学
- Proteomics   蛋白质组学

Case Study: PageRank

☐ An algorithm used by Google to rank websites in their search engine, named after Larry Page, one of Google founders.

谷歌用于在其搜索引擎中对网站进行排名的一种算法，以谷歌创始人之一拉里·佩奇的名字命名。

☐ PageRank works by counting the number and quality of links to a page to determine how important the website is.

PageRank通过计算网页的链接数量和质量来决定该网站的重要性。

☐ The underlying assumption is that more important websites are likely to receive more links from other websites.

其基本假设是：越重要的网站，就会被越多其它网站所链接。

Thank you for your attention !

AI