Introduction to ranking models

Bradley-Terry, Luce, Mallows and CPS

Ling Guang

Department of Computer Science and Engineering,
The Chinese University of Hong Kong

November 1, 2011

1. Introduction
   - Permutations versus rankings
   - Notations
2. Luce
3. Bradley-Terry
4. Mallows
   - Mallows model
   - Distance measure
5. CPS
   - Motivation
   - A tiny bit of group theory
   - Coset-permutation distance
   - CPS model
6. Summary

Permutations versus rankings

Let $\{1, 2, \cdots, n\}$ be a set of objects to be ranked. A ranking π is a bijection from $\{1, 2, \cdots, n\}$ to itself. A ranking have a one-to-one correspondence with a permutation.
Today, we will be talking about probabilistic models on permutations.

Notations

- We use $\pi(i)$ to denote the position given to object i.
- We use $\pi^{-1}(j)$ to denote the object assigned to position j.
- We write π and π^{-1} as vectors whose i-th component is $\pi(i)$ and $\pi^{-1}(i)$, respectively.
- The collection of all permutations of n objects forms a non-abelian group under composition, called the symmetric group of order n, denoted by S_n.

Luce model

Luce model is a stagewise probabilistic model on permutations. It assumes that there is a hidden score $w_i, i = 1, 2, \cdots, n$, for each individual object i. To generate a permutation π, the object $\pi^{-1}(1)$ is assigned to position 1 with probability $\frac{\exp(w_{\pi^{-1}(1)})}{\sum_{i=1}^{n} \exp(w_{\pi^{-1}(i)})}$. Then the object $\pi^{-1}(2)$ is assigned to position 2 with probability $\frac{\exp(w_{\pi^{-1}(2)})}{\sum_{i=2}^{n} \exp(w_{\pi^{-1}(i)})}$. The assignment is continued until a complete permutation is formed.

Luce model

According to the generation procedure described above, the permutation probability of π is,

$$P(\pi) = \prod_{i=1}^{n} \frac{\exp(w_{\pi^{-1}(i)})}{\sum_{j=i}^{n} \exp(w_{\pi^{-1}(j)})}$$

Bradley-Terry model

Like Luce Model, there is a hidden score $w_i, i = 1, 2, \cdots, n$, assigned to each object i. BT model specifies that the probability that item i is ranked higher than j is $\frac{w_i}{w_i + w_j}$. So, Bradley-Terry is a pair-wise model.

Bradley-Terry Model

The probability of observing a full permutation π is

$$P(\pi) = \frac{1}{Z(w)} \prod_{i=\pi^{-1}(1)}^{\pi^{-1}(n-1)} \prod_{j:\pi(j)>\pi(i)} \frac{w_i}{w_i + w_j}.$$

This can be further simplified into

$$P(\pi) = \frac{1}{Z'(w)} \prod_{i=1}^{n} w_i^{n-\pi_i},$$

where $Z'(w)$ is a constant not depending on π.

Mallows model

- Mallows Model is a distribution over permutations parameterized by two parameters
  - Mode: σ
  - Dispersion: θ
- The distribution is defined as

$$P(\pi|\sigma,\theta) = \frac{1}{Z(\sigma,\theta)} \exp(-\theta d(\pi,\sigma))$$

- $d(\pi,\sigma)$ is a distance measure between π and σ and $Z(\sigma,\theta)$ is a normalization constant

Distance measure

There are many well-defined metrics to measure the distance between two permutations.

- Kendall's tau distance

$$d_t(\pi, \sigma) = \sum_{i=1}^{n} \sum_{j>i} 1_{\{\pi\sigma^{-1}(i) > \pi\sigma^{-1}(j)\}}$$

- Spearman's rank correlation

$$d_r(\pi, \sigma) = \sum_{i=1}^{n} (\pi(i) - \sigma(i))^2$$

- Spearman's footrule

$$d_f(\pi, \sigma) = \sum_{i=1}^{n} |\pi(i) - \sigma(i)|$$

Why CPS model

Coset-permutation distance based stagewise (CPS) Model is proposed to tackle the problem with previous ranking models. The expressiveness of Luce model is limited by the specific function of the scores. Yet Luce model is quite efficient. Mallows model, on the other hand, can adapt to many different kinds of distance measure. In this regard, Mallows model have richer expressiveness than Luce model. However, for some of the distance measure mentioned, the normalization constant is very expensive to compute. It requires the summation of $n!$ terms.
CPS is proposed with the hope that it will retain the advantages of both Mallows and Luce while avoiding their limits.

A tiny bit of group theory

Group theory plays a central role in the study of permutations. When we come across more ”advanced” ranking models, like CPS, understanding of basic concepts in group theory becomes inevitable.

What is a group?

In mathematics, a group is an algebraic structure consisting of a set together with an operation that combines any two elements to form a third element. To qualify as a group, the set and the operation must satisfy the following conditions:

- **Closure**: If a and b are in the group then $a \cdot b$ is also in the group.
- **Associativity**: If a, b and c are in the group then $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
- **Identity**: There is an element e of the group such that for any element a of the group:

$$a \cdot e = e \cdot a = a$$

- **Inverses**: For any element a of the group there is an element a^{-1} such that:

$$a \cdot a^{-1} = e$$

$$a^{-1} \cdot a = e$$

Examples of group

- Integer form a group under the operation of addition. The identity element is 0.
- The set of all real numbers except 0 is a group under the operation of multiplication. The identity element is 1.
- The set of one element a is a group with the operation "do nothing". That is, the identity element is a. This is a trivial group.
- The set of all possible permutations of n objects form a group under the operation of composition. The identity element is the identity permutation, i.e. $< 1, 2, 3, \cdots, n >$. This is called the symmetric group of order n, denoted by S_n.

Abelian, infinite group

A group G with an additional property called **commutativity** is **Abelian** group. If it is not commutative, it is a **non-abelian** group.

- **Commutativity** For all a, b in G, we have

$$a \cdot b = b \cdot a$$

A group G is an **infinite** group if the number of elements in the underlying set if infinite, otherwise it is a **finite** group. Generally, finite group is more interesting than infinite group.

Examples of group

- **Abelian, Infinite** Integer form a group under the operation of addition. The identity element is 0.
- **Abelian, Infinite** The set of all real numbers except 0 is a group under the operation of multiplication. The identity element is 1.
- **Abelian, Finite** The set of one element a is a group with the operation "do nothing". That is, the identity element is a. This is a trivial group.
- **Non-Abelian, Infinite** The set of all 2×2 *invertible* matrices under the operation of matrix multiplication. The identity element is identity matrix of size 2×2.
- **Non-Abelian, Finite** The set of all possible permutations of n objects form a group under the operation of composition. The identity element is the identity permutation, i.e. $< 1, 2, 3, \cdots, n >$. This is called the symmetric group of order n, denoted by S_n.

Subgroup

A subset S of a group G is a **subgroup** of G if

- S is closed under the group operation of G.
- The identity element of G is in S.
- For every a in S, a^{-1} is also in S.

Examples of subgroup

- When G is the group of all integers under addition, a (possible) subgroup can be all the numbers divisible by 3 (including 0).
- When G is S_n, S_{n-k} is a subgroup of S_n consisting of all permutations whose first k positions are fixed:

$$S_{n-k} = \{\pi \in S_n | \pi(i) = i, \forall i = 1, 2, \cdots, k\}.$$

Coset

If G is a group, H is a subgroup of G and g is an element of G, then

- $gH = \{g \cdot h | h \in H\}$ is a **left coset** of H in G.
- $Hg = \{h \cdot g | h \in H\}$ is a **right coset** of H in G.

We will be only considering right coset.

Examples of coset

The right coset $S_{n-k}\pi = \{\sigma \cdot \pi | \sigma \in S_{n-k}\}$ is a coset of S_{n-k}. In fact, it denotes all the permutations whose top-k objects are exactly the same as in π.

$$S_{n-k}\pi = \{\sigma | \sigma \in S_n, \sigma^{-1}(i) = \pi^{-1}(i), \forall i = 1, 2, \cdots, k\}$$

Coset-permutation distance

Given a distance measure $d(\sigma, \pi)$ between two permutations π, σ, the **coset-permutation distance** $\hat{d}$ measures the distance between a target permutation σ to a right coset $S_{n-k}\pi$.

Coset-permutation distance is a generalization of the distance between two permutations to the *average* distance between a permutation to a coset. It is defined as

$$\hat{d}(S_{n-k}\pi, \sigma) = \frac{1}{|S_{n-k}\pi|} \sum_{\tau \in S_{n-k}\pi} d(\tau, \sigma).$$

Coset-permutation distance

Actually, coset-permutation distance defines a way to measure the distance between a full rank to a top-k rank.

- The coset $S_{n-k}\pi$ includes all permutations whose top-k objects are the same as π.
- The distance is the average distance between σ and $\tau \in S_{n-k}\pi$.

Coset-permutation distance based stagewise model

Coset-permutation distance based stagewise (CPS) model is similar to Luce model in the sense that they both generate the target permutation in a stagewise manner.

At stage k, the probability of assigning $\pi^{-1}(k)$ to the k-th position is

$$\frac{\exp(-\theta \hat{d}(S_{n-k}\pi, \sigma))}{\sum_{j=k}^{n} \exp(-\theta \hat{d}(S_{n-k}(\pi, k, j), \sigma))}.$$

Coset-permutation distance based stagewise model

$$\frac{\exp(-\theta\hat{d}(S_{n-k}\pi,\sigma))}{\sum_{j=k}^{n}\exp(-\theta\hat{d}(S_{n-k}(\pi,k,j),\sigma))}$$

The $S_{n-k}(\pi,k,j)$ denotes the right coset $S_{n-k}\pi'$ where π' is π with the position of object $\pi^{-1}(k)$ and $\pi^{-1}(j)$ $(j>k)$ switched. That is, $S_{n-k}(\pi,k,j)$ denote all the permutations that rank objects $\pi^{-1}(1),\cdots,\pi^{-1}(k-1)$ and $\pi^{-1}(j)$ in the top-k positions respectively.

Coset-permutation distance based stagewise model

Once defined the probability of selection at stage k, we can now define the probability of obtaining the full ranking π, given dispersion parameter θ and location permutation (mode) σ as

$$P(\pi|\theta,\sigma) = \prod_{k=1}^{n} \frac{\exp(-\theta\hat{d}(S_{n-k}\pi,\sigma))}{\sum_{j=k}^{n}\exp(-\theta\hat{d}(S_{n-k}(\pi,k,j),\sigma))}$$

It is proven that $\hat{d}(S_{n-k}\pi,\sigma)$ can be computed in $O(n^2)$ time for all three distance measure mentioned.

Summary

In this talk, we covered several ranking models.

- Score (Utility) based models
  - Luce model
  - Bradley-Terry model
- Mallows model
- Basic group theory
- CPS

Q&A

Any questions?