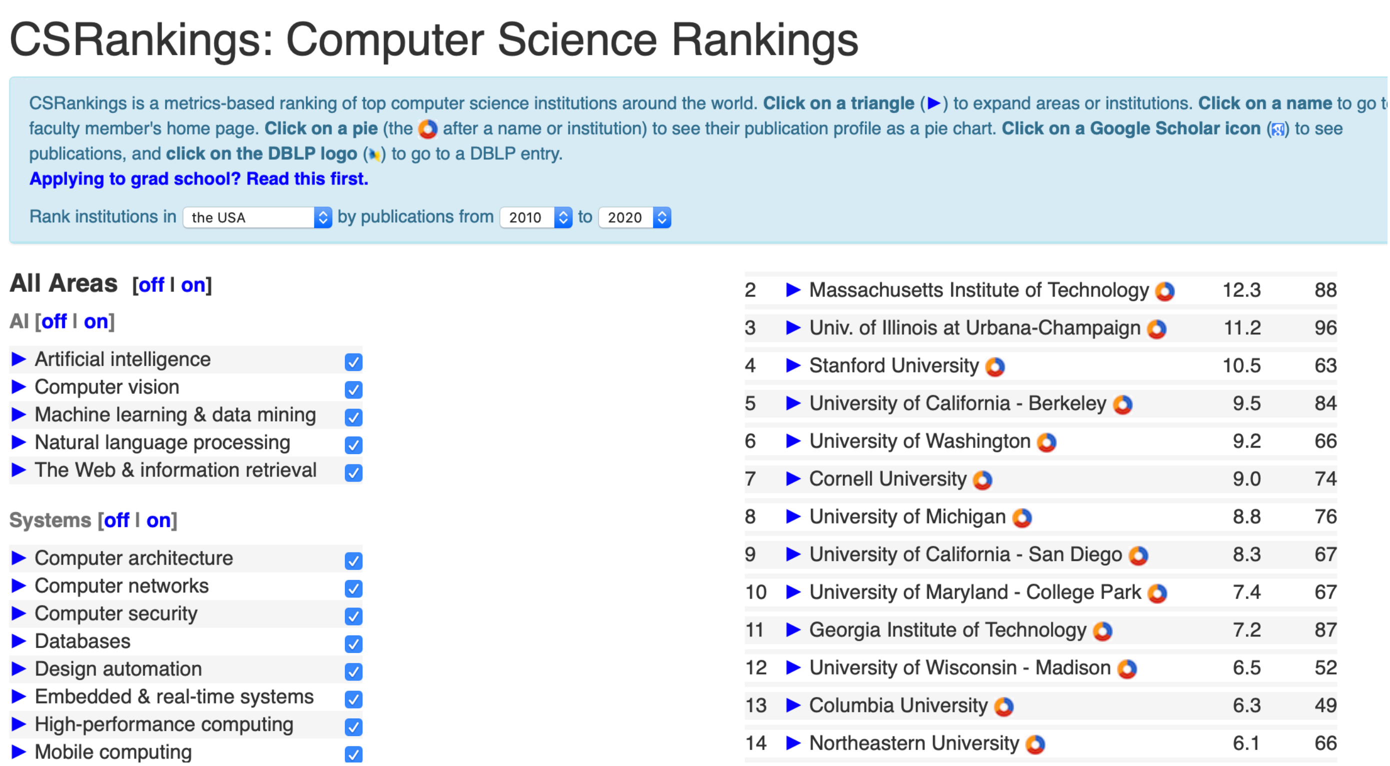
CSI 436/536
Introduction to Machine Learning

LLSE Ranking

Professor Siwei Lyu
Computer Science
University at Albany, State University of New York

Pairwise ranking problem

- Based on paper "A graph interpretation of the least squares ranking method" by Laszalo Csato
- Problem of ranking
  - Get n items and we would like to rank them based on some pairwise comparisons, not all pairs are compared



CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. **Click on a triangle** (▶) to expand areas or institutions. **Click on a name** to go t faculty member's home page. **Click on a pie** (the after a name or institution) to see their publication profile as a pie chart. **Click on a Google Scholar icon** to see publications, and **click on the DBLP logo** to go to a DBLP entry.

Applying to grad school? Read this first.

Rank institutions in the USA by publications from 2010 to 2020

All Areas [off | on]

AI [off | on]

- ▶ Artificial intelligence
- ▶ Computer vision
- ▶ Machine learning & data mining
- ▶ Natural language processing
- ▶ The Web & information retrieval

Systems [off | on]

- ▶ Computer architecture
- ▶ Computer networks
- ▶ Computer security
- ▶ Databases
- ▶ Design automation
- ▶ Embedded & real-time systems
- ▶ High-performance computing
- ▶ Mobile computing

2	▶ Massachusetts Institute of Technology	12.3	88
3	▶ Univ. of Illinois at Urbana-Champaign	11.2	96
4	▶ Stanford University	10.5	63
5	▶ University of California - Berkeley	9.5	84
6	▶ University of Washington	9.2	66
7	▶ Cornell University	9.0	74
8	▶ University of Michigan	8.8	76
9	▶ University of California - San Diego	8.3	67
10	▶ University of Maryland - College Park	7.4	67
11	▶ Georgia Institute of Technology	7.2	87
12	▶ University of Wisconsin - Madison	6.5	52
13	▶ Columbia University	6.3	49
14	▶ Northeastern University	6.1	66

Problem setting

- Objective function $\min_r \sum_{ij} m_{ij}(r_i - r_j - q_{ij})^2$
  - $q_{ij} = -q_{ji}$, for items i and j, as their comparative scores, we denote the set of all such pairs as S.
  - $m_{ij} = 1$ if $(i,j) \in S$, and 0 if $(i,j) \notin S$.
    - If there is no comparison, don't care the error
  - r_i is the rank of the ith item
  - We denote $M_{ij} = qij$ if $(i,j) \in S$, and 0 if $(i,j) \notin S$.
    - Matrix M is anti-symmetric, i.e, $M^T = -M$

Derivation

- Expand the objective function
  $$\sum_{ij} m_{ij}(r_i - r_j - q_{ij})^2 = \sum_{ij} m_{ij}(r_i - r_j)^2 - 2\sum_{ij} M_{ij}(r_i - r_j)$$
  - First term
    $$\sum_{i,j} m_{ij}(r_i - r_j)^2 = \sum_{ij} m_{ij}r_i^2 - 2\sum_{ij} m_{ij}r_ir_j + \sum_{ij} m_{ij}r_j^2 = 2\sum_i r_i^2 \sum_j m_{ij} - 2\sum_{ij} m_{ij}r_ir_j$$
    - Introduce a diagonal matrix D with $D_{ii} = \sum_j m_{ij}$
    - Matrix $A_{ij} = m_{ij}$
    - Then this becomes $2r^T(D - A)r$
  - The second term (using anti-symmetry of M)
    $$2\sum_{ij} M_{ij}(r_i - r_j) = 2\sum_{ij} M_{ij}r_i - 2\sum_{ij} M_{ij}r_j = 2(1^TM^Tr - 1^TMr) = 4 \cdot 1^TM^Tr$$
- The objective function becomes $2r^T(D - A)r - 4 \cdot 1^TM^Tr$

LLSE ranking algorithm

- Minimizing $2r^T(D-A)r - 4 \cdot 1^T M^T r$, ignoring constants, we get solution given by $(D-A)r - M1 = 0$
- The relative ranking vector r is given by solving $(D-A)r = M1$
- We can derive a linear ranking function $f(x) = w^T x$, the corresponding problem becomes $\min_w \sum_{ij} m_{ij}(w^T x_i - w^T x_j - q_{ij})^2$
- The vector version of the objective is then $\min_w 2w^T X(D-A)X^T w - 4w^T XM1$, and the solution is given by $X(D-A)X^T w = XM1$
    - This is known as LLSE ranking solution

Graph interpretation

- Construct a graph G with each data point a node
- If there is a comparison between node (i,j) then we put a pair of directed edges between them
- The weights on the edges are given by q_{ij}
  - Matrix A is the **adjacency matrix** of this graph
  - Every weighted undirected graph is determined uniquely by a matrix
- Matrix L = D-A is the **graph Laplacian** of G
- There is an intimate relation between graph theory and linear algebra
  - We seem more of this for spectral clustering