Aggregating Quantitative Relative Judgments: From Social Choice to Ranking Prediction

NEURAL INFORMATION PROCESSING SYSTEMS

Yixuan Even Xu — Carnegie Mellon University
Hanrui Zhang — Chinese University of Hong Kong
Yu Cheng — Brown University
Vincent Conitzer — Carnegie Mellon University

Ranking Aggregation

Problem: Aggregating multiple input rankings into an integrated one

The problem is of interest in multiple research communities

- **Voting theory:** each **voter** ranks the candidates, and a **voting rule** decides a winning candidate or a ranking of all candidates
- **Learning-to-rank:** ranking **web pages** in response to a search query, or ranking **recommendations** to a user
- **Common ground:** there is a latent "true" ranking of the elements, of which all inputs are just noisy observations

Focuses of Different Communities

Voting theory (social choice)
- **Inputs: subjective**
- **Desiderata:**
  - **transparency**
  - **simple voting rule**
  - **strategy-proofness**

This Work

Learning-to-rank
- **Inputs: objective**
- **Desiderata:**
  - **relevance to query**
  - **recomendation quality**

This work: An attempt to bridge the two communities by applying formulations from social choice to learning-to-rank problems and inputs

Judgment Aggregation

Quantitative judgment aggregation:

- A way to think of ranking aggregation **in social choice**
- **Aggregation inputs:** Quantitative relative judgments $\{(a, b, y)\}$, i.e., "Candidate a is better than candidate b by y units quantitatively"
- **Example in social choice:** "Using 1 unit of gasoline is 3 times as bad as creating 1 unit of landfill trash" (in a societal tradeoff context)

We observe that the relative "judgments" can also be produced by an **objective process** other than a subjective agent reporting

- Applying judgment aggregation **formulations from social choice**
- To **learning-to-rank instances** (so quality overweighs other social choice desiderata like strategy-proofness)
- This conceptually bridges the two communities

Example Application: Races

Races are one example of objective judgments

- **Judgment:** "Alice is faster than Bob by 11 minutes in Boston Marathon"
- **Aggregation result:** An ordering of the contestants' strengths

Concrete examples of the problem & Why we need more than mean / median

	Boston	New York	Chicago
Alice	4:00:00	4:10:00	3:50:00
Bob	4:11:00	4:18:00	4:01:00
Charlie	N/A	N/A	4:09:00

- Bob seems to be faster than Charlie judging from the Chicago race
- But mean / median draws the opposite conclusion

	Boston	New York	Chicago
Alice	N/A	4:10:00	N/A
Bob	4:11:00	4:18:00	4:01:00
Charlie	N/A	N/A	4:09:00

- Alice was faster than Bob in New York, and Bob was faster than Charlie in Chicago, thus, we can deduce that Alice seems to be faster than Charlie
- But this is not possible if we only look that Alice's and Charlie's histories

QRJA Aggregation Rule

Quantitative relative judgment aggregation (QRJA) rule: Given a set of m quantitative relative judgments $\{(a_i, b_i, y_i)\}$ about n contestants, and the weights $\{w_i\}$ of these judgments, find a vector $\boldsymbol{x} \in R^n$ that minimizes

$$\sum_{i=1}^{m} w_i \cdot f\left(\left|(x_{a_i} - x_{b_i}) - y_i\right|\right)$$

$f: R \to R$ maps the inconsistency with judgements to loss

- $\boldsymbol{f(x) = x}$**:** prior work (Conitzer et al., 2016, Zhang et al., 2019)
- **If** $\boldsymbol{f(x)}$ **is convex:** solvable in polynomial time
- $\boldsymbol{f(x) = x^p}$**: The focus of this work**

Computational Complexity

We provide **a tight characterization** of ℓ_p QRJA's complexity

- When $p \geq 1$, ℓ_p QRJA can be solved in **almost-linear** time $O(m^{1+o(1)})$
  - Technique(s): Lagrangian dual, reduction to network flow
- When $p < 1$, ℓ_p QRJA is **NP-Hard**, and there is no FPTAS
  - Technique(s): reduction from Max-Cut

Additionally, we show that when $p \in [1, 2]$ and $m \gg n$, we can **reduce** $\boldsymbol{m}$ **to** $\boldsymbol{\tilde{O}(n)}$ while incurring a small error

- Technique(s): subsampling input judgments according to Lewis weights

Experiments

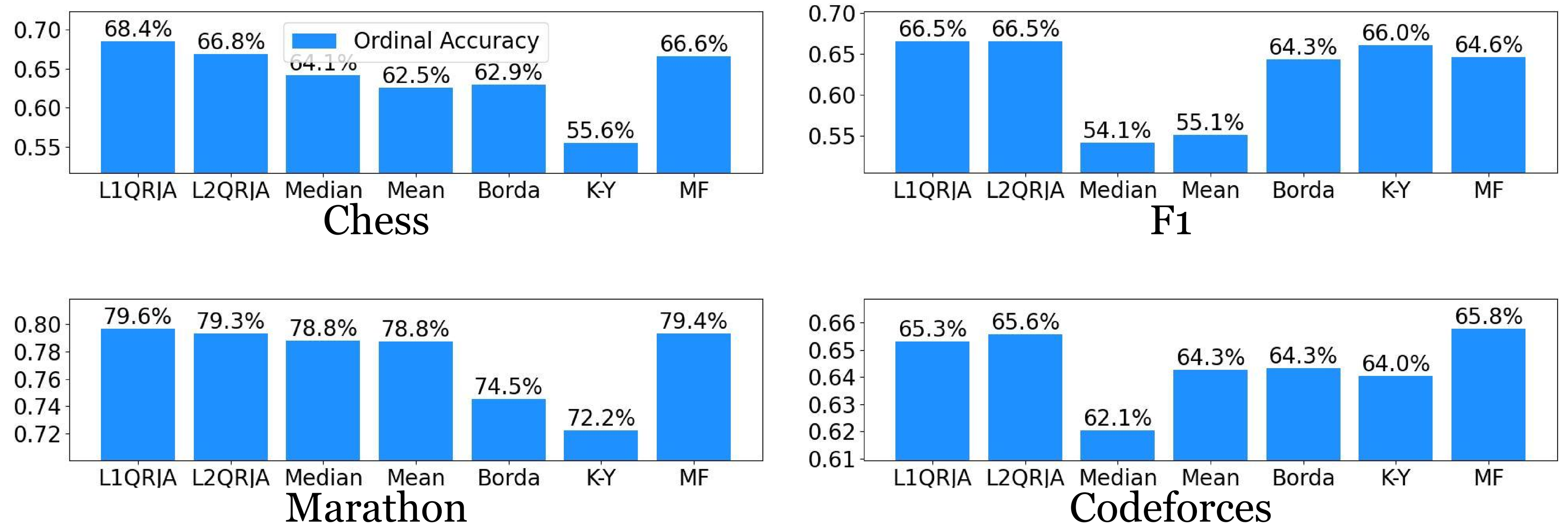
We conduct experiments on real-world race data

- **Datasets:** F1 races, marathon, programming contests, chess, etc.
- **Our algorithms:** ℓ_1 and ℓ_2 QRJA

Benchmarks

- **Simple** benchmarks: Mean, Median
- From **social choice:** Borda, Kemeny-Young
- From **learning-to-rank:** Matrix Factorization

We look at **ordinal accuracy (shown below)** and **quantitative loss**



Both MF and QRJA are never significantly worse than the best algorithm on any of the tested datasets, and QRJA additionally offers an interpretable model