Overview

The modern online marketplace relies on reputation systems, but these systems are often plagued with problems stemming from pressure and various cognitive biases, leading to reputation inflation and noisy ratings. In this work, we argue that marketplaces should adopt a pairwise-comparison based reputation system. We characterize convergence of reputation systems to a market equilibrium and study the resulting fixed point joint distribution of true and estimated ratings.

Key Questions:
- Under what market conditions does a reputation system achieve a steady state?
- At steady state, do pairwise comparisons better recover the true ranking than cardinal systems?

Background

Reputation Systems

Problems of cardinal systems
- Inflation: On eBay, >90% of sellers had >98% positive rating [NT15]
- De facto binarization [HZP09]
- Little marginal information per rating

Common Fixes
- Simultaneous reveal
- Private and anonymous ratings
- Implied ratings from behavior
- Internalize rating externality [G16]

Pairwise Comparisons

Long History
- Voting: Condorcet
- Chess: Elo
- Xbox Live: Trueskill

Varied Approaches
- Model based recovery
- Axiomatic
- Graph based
- Iterative

Work in noisy settings non-i.i.d sampling of pairs

Model

Buyers B, Sellers S Mass μb, μs = 1 of Buyers, Sellers. Seller true quality θ ∈ [0, 1] ~ w(θ), estimation state x ∈ X. Each time k, mass λs of sellers is reborn.

Object of Study

Evolution of the joint density, μk(θ, x).

Matching and Search Rate

At time k, each buyer matches i.i.d with a seller at rate g(x)μk(x), where g is a matching function such that ∫X g(x)μk(x)dx = 1 and g(x)μk(x) ≤ 1 ∀x ∈ X, μk(x).

Feedback

Buyer observes transaction quality, gives binary feedback, positive w.p. β(θ).

Cardinal β(θ) = min(θ + ε, 1), where ε captures any inflationary pressure.

Landmark Comparisons

Buyer compares seller with landmark comparative experience with quality x∗ ~ w∗ described by the platform. β(θ|x∗) = 1[θ + ε ≥ x∗].

Then β(θ) = Mε(θ + ε).

State Update

fkw(x|ξ, x∗) = state update functions after a win and loss.

Number of wins X = N2, x0 = (0, 0). Then, fkw(x|j, ξ, x∗) = δ(x = j + 1, ξ + 1) and fkw(x|j, ξ, x∗) = δ(x = j, ξ + 1).

Linearized Elo

X = R, x0 = 0, K1, K2 ∈ R. Then,

fK1, K2(ξ, x∗) = δ(k1 + x∗(1 − 1/K2) + x∗/K2)

Dynamical System

μk+1(θ, x) = M(θ = x0)w(θ) + (1 − λ)∫(1 − g(ξ)μk(ξ, x))μk(θ, x)

+ ∫X μk(θ, x)g(ξ)μk(x, ξ)∫x∗ Mε(θ|x∗)fkw(x|ξ, x∗) w∗(x∗)dx∗ dξ

Theory

Under what market conditions and update rules does the dynamical system converge? Are such conditions typical?

Recovery of True Ranking

Thm. Under general conditions, rankings from landmark comparisons have a higher Kendall’s Tau rank correlation with the true ranking than rankings from cardinal ratings.

Why? Recover more marginal information from each rating

Intuition

Suppose seller quality is skewed:

With cardinal ratings, probability of a good rating is equally skewed.

With landmark comparisons, probability of winning can be designed:

Real-world Data

Used implied pairwise comparisons on 3 real world platforms:
- Freelancer hiring on popular online platform
- Course selections by undergraduates
- Movie ratings from MovieLens

Pairwise comparisons better predicted future ratings, and ratings were well-behaved.

Conclusion

We develop a general model for reputation systems on online platforms and characterize the conditions under which the system is well behaved.

We then show that under this model in general market conditions, pairwise comparison based systems better recover the true underlying ranking.

Finally, we leverage data from 3 platforms to demonstrate real-world performance.