Unfair rating detection in Reputation management systems

Long Ngo
Agenda

- Reputation management systems
- Unfair rating problem
- Unfair rating detection
Agenda

• **Reputation management systems**
• Unfair rating problem
• Unfair rating detection
Reputation in online communities (1)

- Virtual identity
- Easy to join
- Remote
- Little control
Reputation in online communities (2)

- Virtual identity
- Easy to join
- Remote
- Little control
- Reputation
Reputation in online communities (2)

- Virtual identity
- Easy to join
- Remote
- Little control
- Reputation: in \([0,1]\)
Agenda

• Reputation management systems
• **Unfair rating problem**
• Unfair rating detection
3 unfair rating types

• Ballot-stuffing:
  – Say too good about a seller
  – Always give 1

• Bad-mouthing
  – Say too bad about a seller
  – Always give 0

• Complementary
  – Always give opposite: 0->1, 1->0
Agenda

• Reputation management systems
• Unfair rating problem
• **Unfair rating detection: 7 methods**
2 types of methods

• Endogenous
  – Just use the set of ratings
  – Statistically analyse

• Exogenous
  – Use the set of ratings
  – And also other information:
    • Habit, IP address, and so on
Agenda

• Reputation management systems
• Unfair rating problem
• **Unfair rating detection: 7 methods**
  – Endogenous: 4 methods
1.1: Comparing ratings

- **Input**: a set of ratings
  - **Output**: who are fair
- Compute the total reputation from the set
- Compare each one’s ratings with the total one
  - **Unfair if too different**
- Repeat until no change
1.1: Comparing ratings

- Outside of the bound $\rightarrow$ Too different
1.2: Comparing entropy

- Pretty similar to the previous one
- But compare Entropies of ratings
- Example:
  - A set of ratings: 4 positive, 1 negative
  - Probability of positive again:
    - \( P_p = \frac{4}{4+1} = 0.8; \ P_n = \frac{1}{4+1} = 0.2 \)
  - Entropy \( H = - P_p \log(P_p) - P_n \log(P_n) = 0.217 \)
  - If the total says "my H is 0.8" -> unfair set
1.3: Using reputation tree

- Evaluate quality of a rater by his ratings on everything
- Use that quality as the weight value to compute reputation of something
1.3: How to evaluate raters

- Quality of a rating: depends on how much it is similar to others’ ratings
1.3: How to evaluate raters

- Quality of a rating: depends on how much it is similar to others’ ratings
- Then just combine upwards
4: Controlled anonymity
4: Controlled anonymity

- ID1: 80% honest
- ID2: 80% honest
- ID3: 80% honest
- ID4: 80% honest

• No more bad-mouthing
1.4: Clustering

- Because no bad-mouthing

  - 40% good
  - 43% good

Shh!, It’s me

90% good

92% good

89% good
1.4: Clustering

- Guaranteed: No bad-mouthing
  - 40% good
  - 43% good

- Now: No ballot-stuffing
  - Bad group
    - 90% good
    - 92% good
    - 89% good
Agenda

• Reputation management systems
• Unfair rating problem
• **Unfair rating detection: 7 methods**
  – Endogenous: 4 methods
  – Exogenous: 3 methods
2.1 Clustering and IP checking

- To detect if a node A in P2P creates unreal nodes for supporting him
- Clustering like the previous method
- Take random nodes in a cluster then check IP. If too many do not exist, then the node A is deceptive
2.2 Using trust rating

Blue is good

Blue is bad

Red is bad
2.2 Using trust rating

Many say: Red is bad
2.3: Weighted Majority Algorithm

- First round

I trust you equally
2.3: Weighted Majority Algorithm

- After a transaction

- Oh, Blue is good. I will trust Purple less
Time to conclude?

• Few remarks:
  – Reputation change step by step
  – Sometimes order of checking ratings affect the result, but little
  – No solution is always good or always bad in all cases

• Still much of future work, but people are happy with eBay
Thank you

Any easy question?