Inferring Player Rating from Performance Data in Massively Multiplayer Online Role-Playing Games (MMORPGs)

Kyong Jin Shim

University of Minnesota – Twin Cities
Data Mining Research Lab

kjshim@cs.umn.edu

Last updated: 08/21/2009

Outline

- Introduction (.5 min)
- Motivation (1 min)
- Performance metrics (1 min)
- Impact of Group Formation (1 min)
- Methods (5 min)
- Results (11~ min)
- Future Directions (2 min)
- Q & A (10 min)

Introduction

- Funding project: Virtual World Exploratorium (Sony Online Entertainment's EverQuest II)
- Conference accepted
  - IEEE Social Computing - Workshop on Social Intelligence in Applied Gaming (SIAG09)
- Collaborators
  - Professor Jaideep Srivastava (faculty, research)
  - Muhammad Ahmad (research)
  - Nishith Pathak (research)
- Acknowledgements
  - DMR Lab: Jaya Kawale (statistical analysis questions, tools)
  - USC, Northwestern, U of Illinois for helpful discussions.

Motivation

- Educational Psychology, Learning Sciences
  - Learning = Trajectory of participation (Greeno, 1997)
  - Learning = Growth of identity within community (Gee, 1999)
  - Learning
    - How individual interacts with
      - Materials
      - Social contexts
    - How interactions change over time (i.e. grouping behavior)
    - How individual constructs knowledge (i.e. apprenticeship, completing tasks)
  - Learning → Transformation of individual from legitimate peripheral participant (Lave & Wenger, 1991) to <u>central member</u> of a community.

Virtual Environments

- Sony Online Entertainment's EverQuest II
  - Massively Multiplayer Online Role-Playing Game
    - Players sign on simultaneously to the same online, persistent virtual world
    - Interact and collaborate with each other through in-game characters
- Venue for studying:
  - Individual/group learning behaviors and trajectories.
  - Social interactions, literacy, community membership.
- Social interactions are easier to capture compared to the real world.
  - Today's computing technologies allow for speedy data capturing.
  - Apply data mining techniques for analysis.

Performance Metrics

- Long been studied in Industrial Engineering – Operations Research.
- Performance = Productivity + Quality + Inventory
- Performance Metrics
  - Assembly line balancing problem: maximize efficiency through minimization of idle time.
  - → Maximum possible productivity in a given time duration.
  - Can we leverage this for measuring online player's performance?
  - Being able to measure player's performance over time across difficulty levels → Allow for individual/group learning patterns

Impact of Groups on Performance

- Operations Research
  - Recent trend in manufacturing plants to adopt formation of work teams as a practice.
  - Goal: Increase performance, especially quality.
- MMORPGs
  - Nature of games encourage group formations (quests).
  - Homogeneous vs. Heterogeneous group formations.
    - Monster raids vs. Quests
  - How does group formation affect individual player performance?

Performance Matrix

	T1 (L = 5, SL = 2)	T2 (L = 6, SL = 10)	T3 (L = 7, SL = 1)	T4 (L = 9, SL = 2)	T5 (L = 11, SL = 8)
P1 (L = 30)	80 1/1	126 ½	135 1/4	173 1/4	199 1/10
P2 (L = 25)	140 1/2	200 1/3	x	320 1/3	273 1/16
P3 (L = 25)	140 1/4	200 1/6	260 1/8	x	x
P4 (L = 15)	250 1/10	357 1/10	400 1/16	x	x

Experience Points

Success Ratio

T = task (monster). Each monster has 1) Level and 2) Sub-level
P = player. Each player has a Level (i.e. L = 25)

Objective – Player Rating Methods

- How does an individual learn in EverQuest II? How does he progress through levels in the game?
  - Progress = indication of learning
  - Analyze performance data
  - Performance = f(Productivity, Quality)
  - Completing a set of tasks
    - Selection of tasks (i.e. choosing easy tasks vs. choosing challenging tasks)
    - Duration of task completion, number of attempts
    - Task difficulty, points gained
  - Compute player ratings.
- Past performance = predictor of future performance?

Performance Metrics Index #1
Performance = F (Productivity)

- Productivity
  - Number of tasks completed
  - <u>Amount of Experience points gained</u>
  - Time factor: Total actual play time (session time)
    - Player can be signed on for 24 hours, but the actual play time to complete one or more tasks could be less than that.
    - Not all players are task-oriented, meaning they could spend time socializing with other people, not necessarily completing tasks.

Performance Metrics Index #1
Performance = F (Productivity)

$$Performance_k = \frac{\sum_{i=1}^{N} XP_i}{\sum_{j=1}^{M} ST_j}$$

where

XP = Experience points

N = Total number of tasks completed by Player K

ST = Session time

M = Total number of session times during which Player K completed tasks

- Performance of player K at a certain Level

Performance Metrics Index #2
Performance = F (Productivity, Quality)

- Quality
  - How well can a given player complete one or more tasks?
  - <u>Success Ratio</u>
    - (# successful attempts) / (# successful attempts + # unsuccessful attempts)

Performance Metrics Index #2
Performance = F (Productivity, Quality)

$$Performance_k = \frac{\sum_{i=1}^{N} \frac{XP_i + (XP_i \times Q_i)}{2}}{\sum_{j=1}^{M} ST_j}$$

where

XP = Experience points

N = Total number of tasks completed by Player K

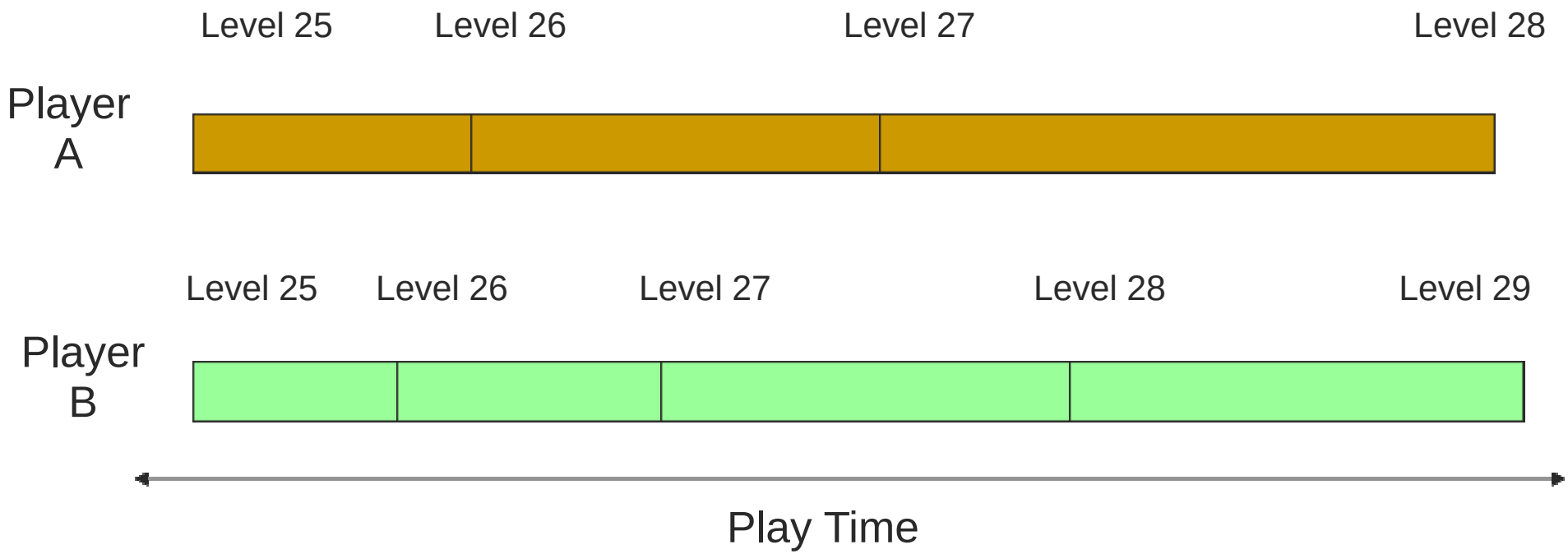
ST = Session time

M = Total number of session times during which Player K completed tasks

Q = Quality or success ratio associated with completing Task i

Is Performance Predictable?

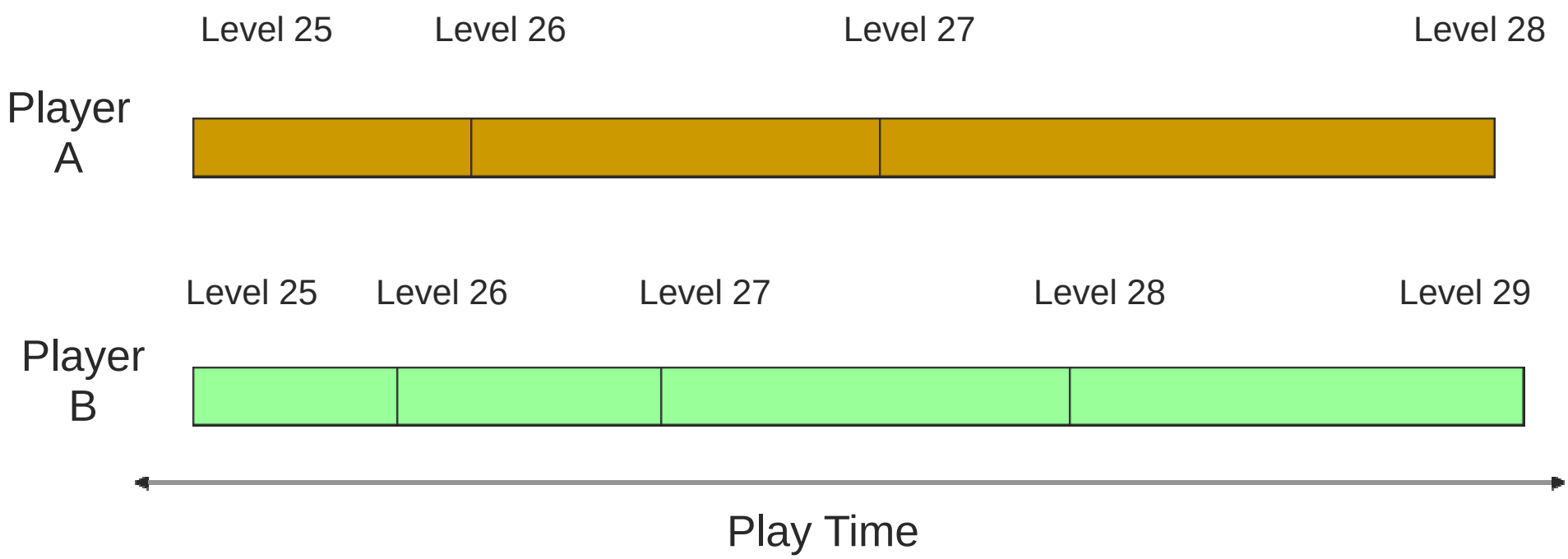
- Is Past performance a good predictor of a player's future performance?



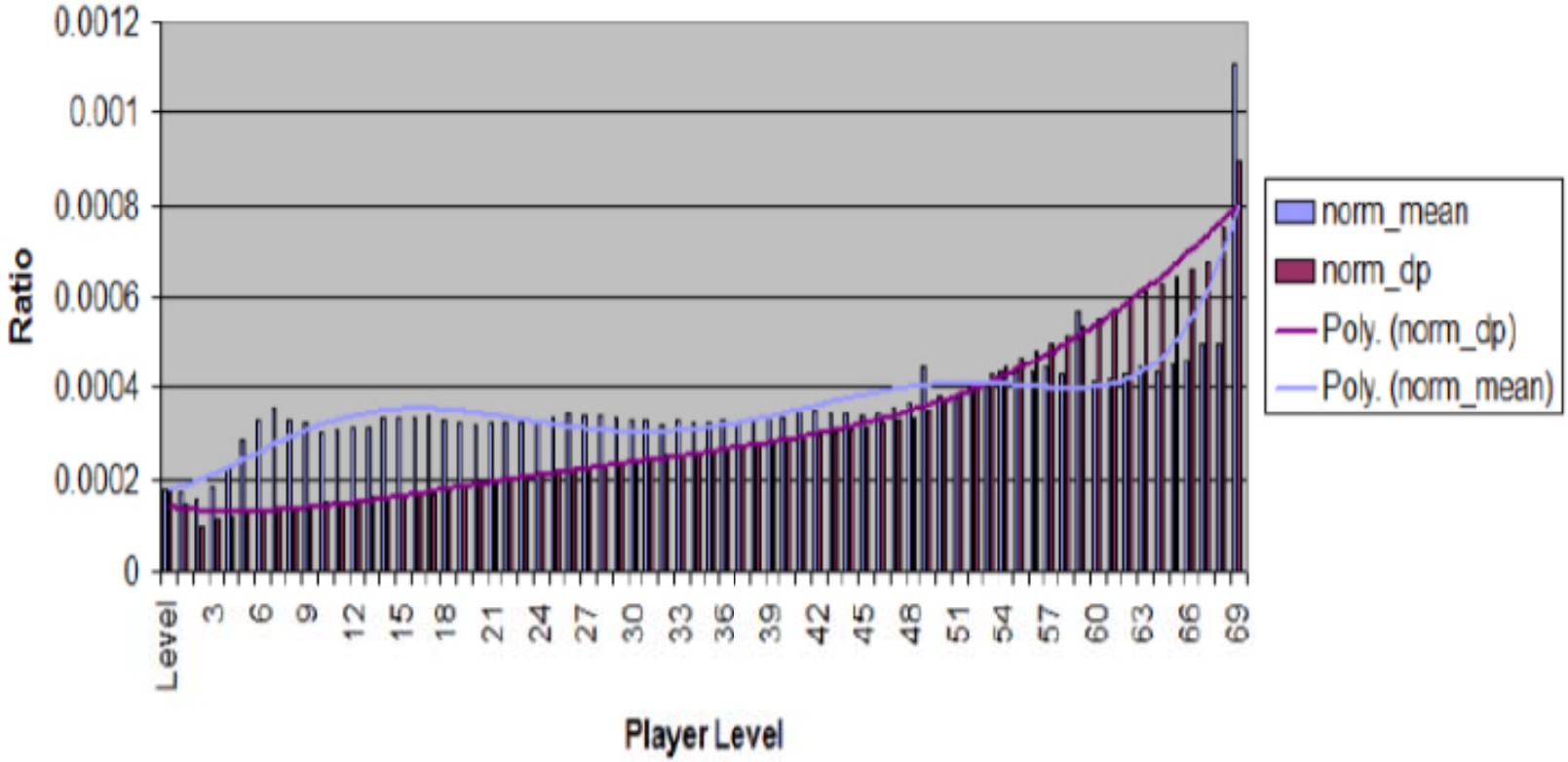
Per definition of Performance Metrics methods #1 and #2, Player B is a better player.
Player B took less time to advance from Level 25 to Level 26.
Consistently Player B shows that he is a better player than Player A.

Is Performance Predictable?

- Catch
  - Past is a good predictor of future performance. But, how far back do we go?
  - Is Player A's performance between Level 25 and Level 26 a good predictor of his performance down the road, say to advance to Level 70?



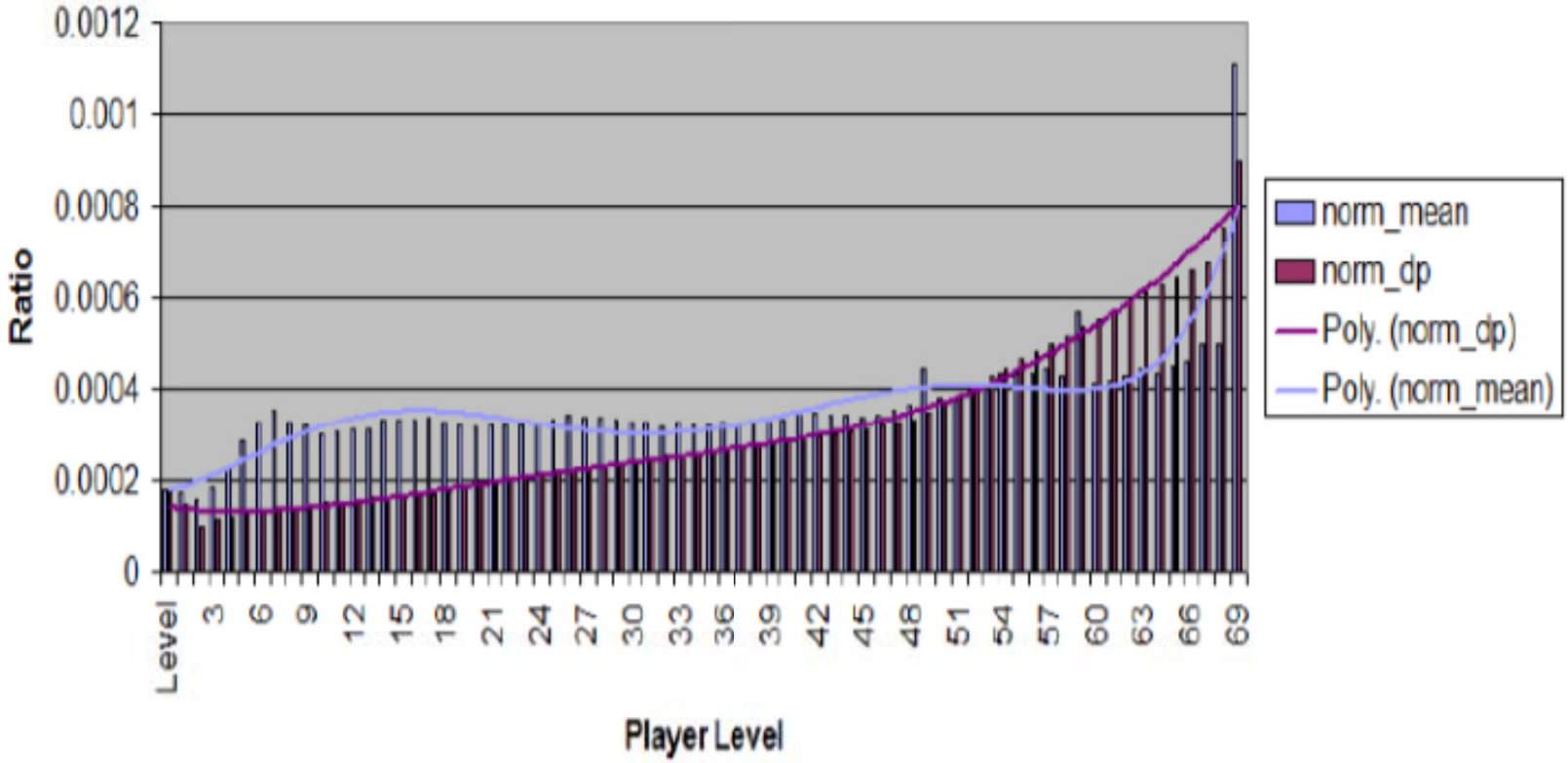
Performance data reveals…



- Dataset: March, 2006 ~ August, 2006 XP data (Monster Kills ONLY)
- The game's ding points are indicative of player level difficulty or how much effort is needed to move from Level i to Level i + 1.

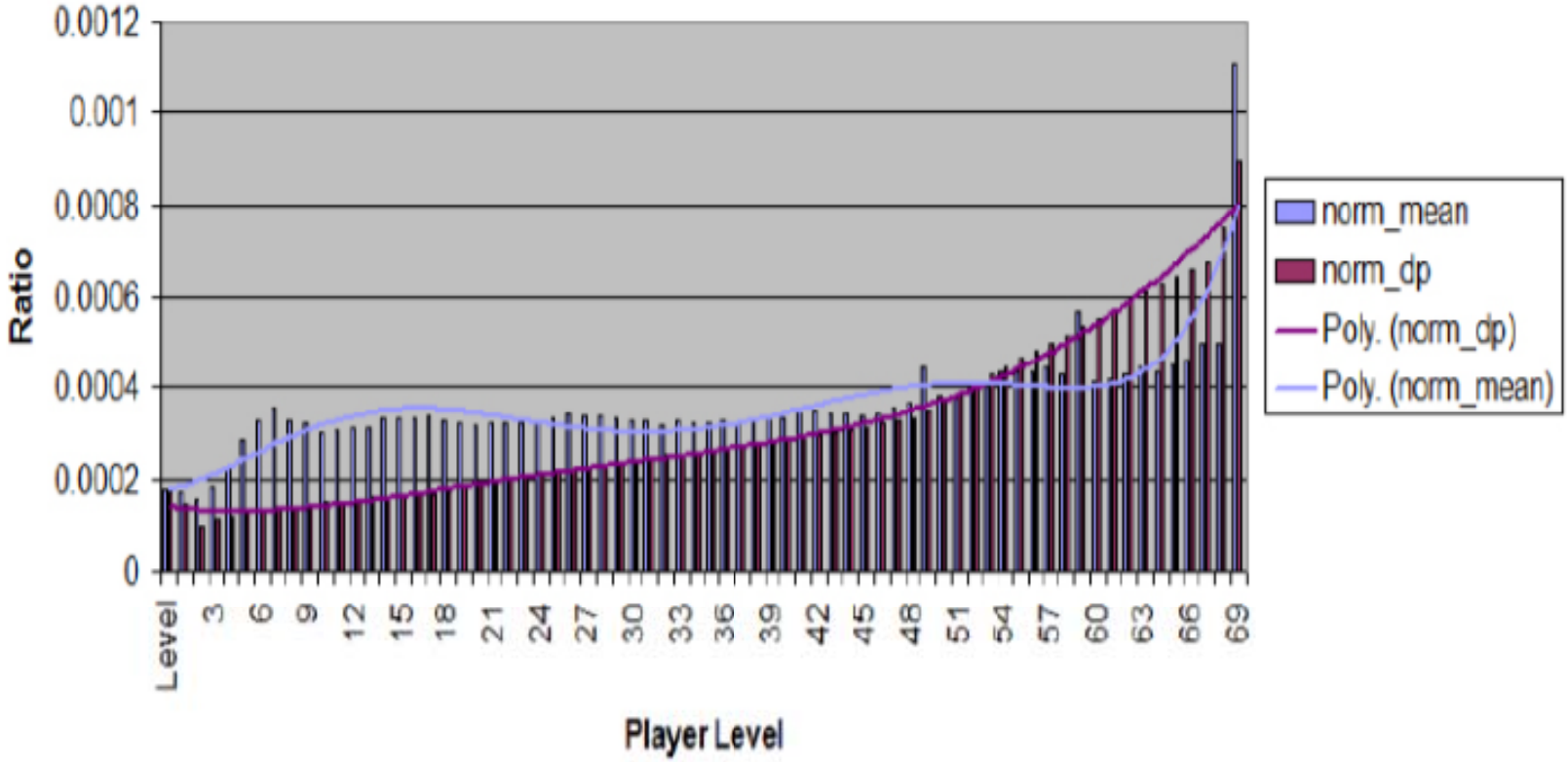
- Another source of player level difficulty is the game's performance data.
- How much effort is actually being spent to move from Level i to Level i + 1

Performance data reveals… (findings)



- Tasks performed by players up until Level 49 were more challenging than expected as time spent increases with an increasing level of task difficulty.
- Between Levels 50 and 55, the actual time spent is well in accordance with what is expected.
- Beyond Level 55 up until Level 68, the actual time spent is well below what is expected.
- Tasks performed were not challenging enough as time spent decreases with a decreasing level of task difficulty.

Performance data reveals… (applications)



- Understanding of challenge associated with EverQuest II (keeping players entertained through challenges while not pushing them too much)
- Construction of player training programs (i.e. games, military, etc.)

Impact of Group Formation on Individual Player Performance

Fig. 3. Success Ratio - Playing Solo vs. Playing in Groups

- Is it better to play alone or as part of a group? Is the success ratio higher when playing alone vs. playing in groups?
- Dataset: March, 2006 XP data (Monster Kills ONLY)

Impact of Group Formation on Individual Player Performance (findings)

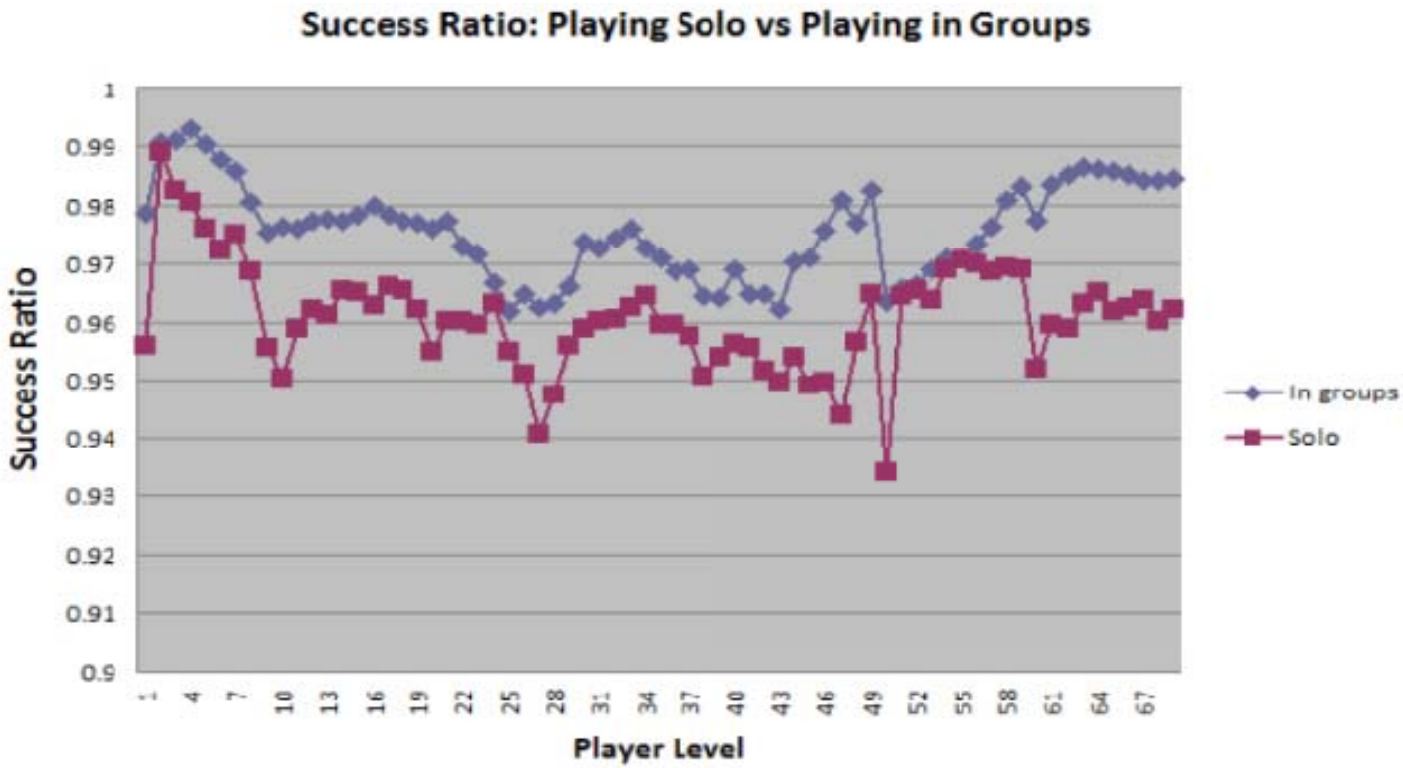


Fig. 3. Success Ratio - Playing Solo vs. Playing in Groups

- In most levels, individual players' success ratios are higher when they played in groups.

Impact of Group Formation on Individual Player Performance (findings)

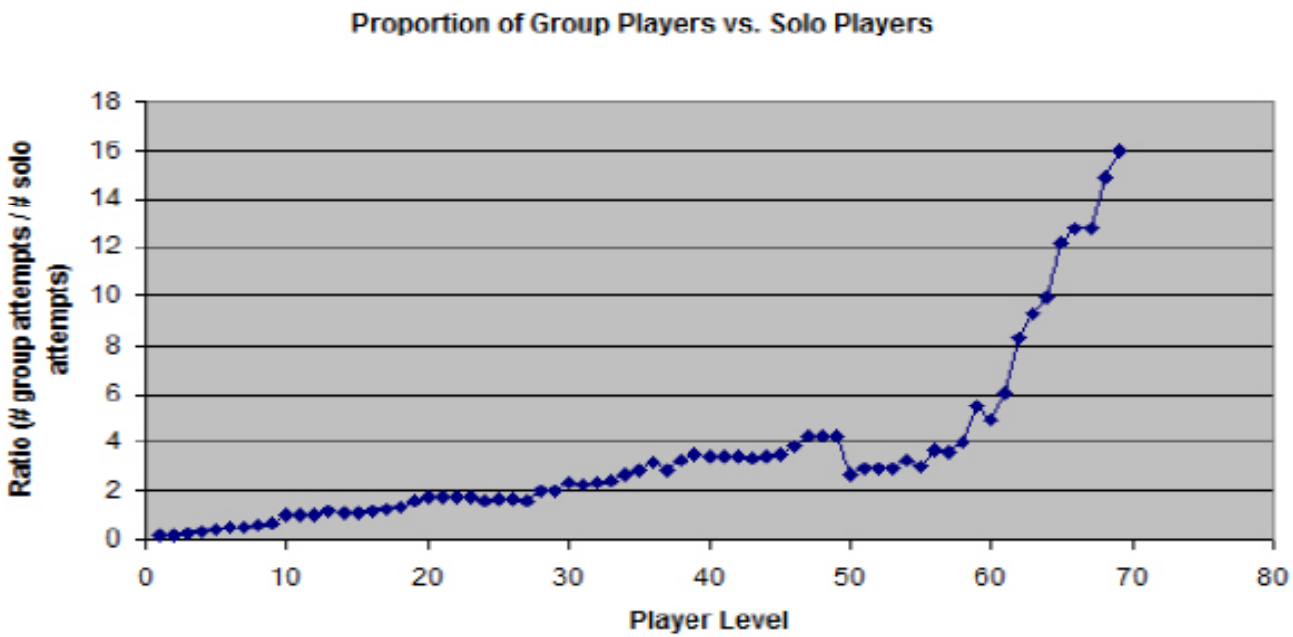


Fig. 4. Proportion of Group Players versus Solo Players

- As the player level increases, the proportion of players playing in groups increases.
- Higher level players tend to fight more difficult monsters which presents the necessity to group with other players in order to successfully kill the monsters.

- Additionally, players reaching higher levels are more inclined to join guilds or raid groups. Moreover, a player's higher level status attracts other players to group with him.

Performance Metric 1 - Evaluation

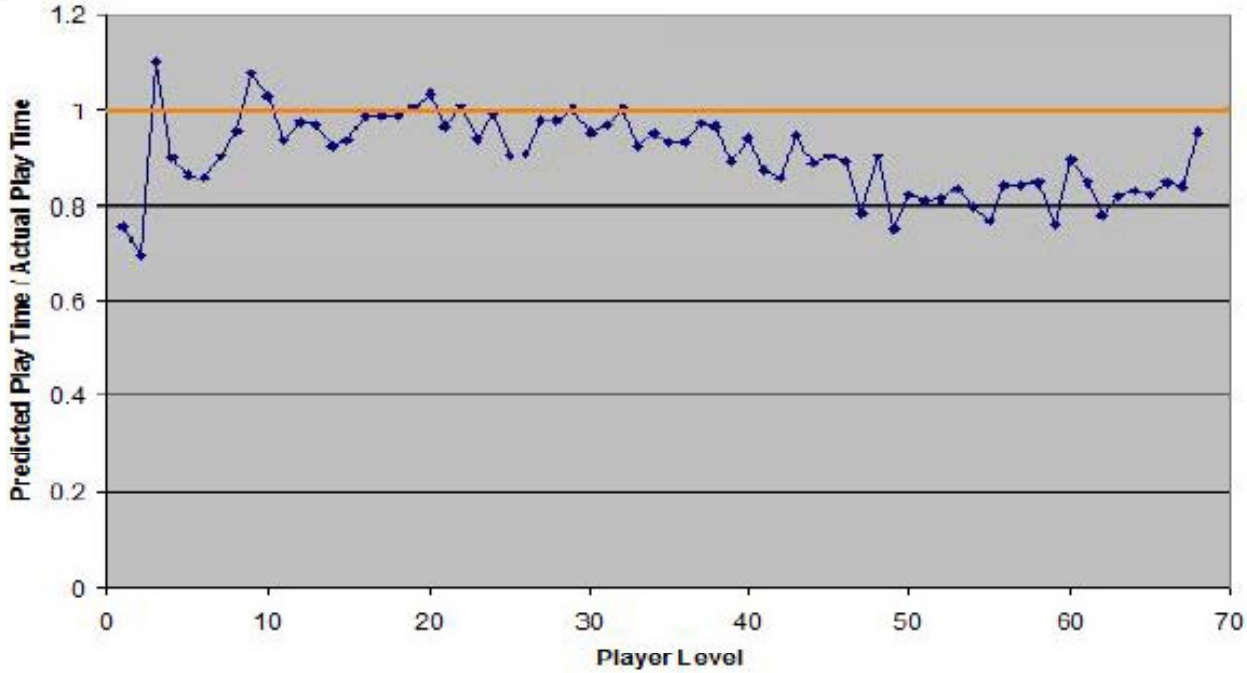


Fig. 5. Ratio of Predicted Play Time and Actual Play Time by Player Level

- At each player level (Level i), we select N players and compute their Performance Metric 1 scores.
- The game's existing ding points-based point scaling system dictates that there is

a fixed amount of points to be gained between any Level i and Level i+1.

- Given a player's Performance Metric 1 score and the fixed amount of points between Level i+1 and Level i + 2, we can compute the total session time (play time) and this becomes our prediction as to how fast this player will advance to the next higher level in the future.

Performance Metric 1 – Evaluation (findings)

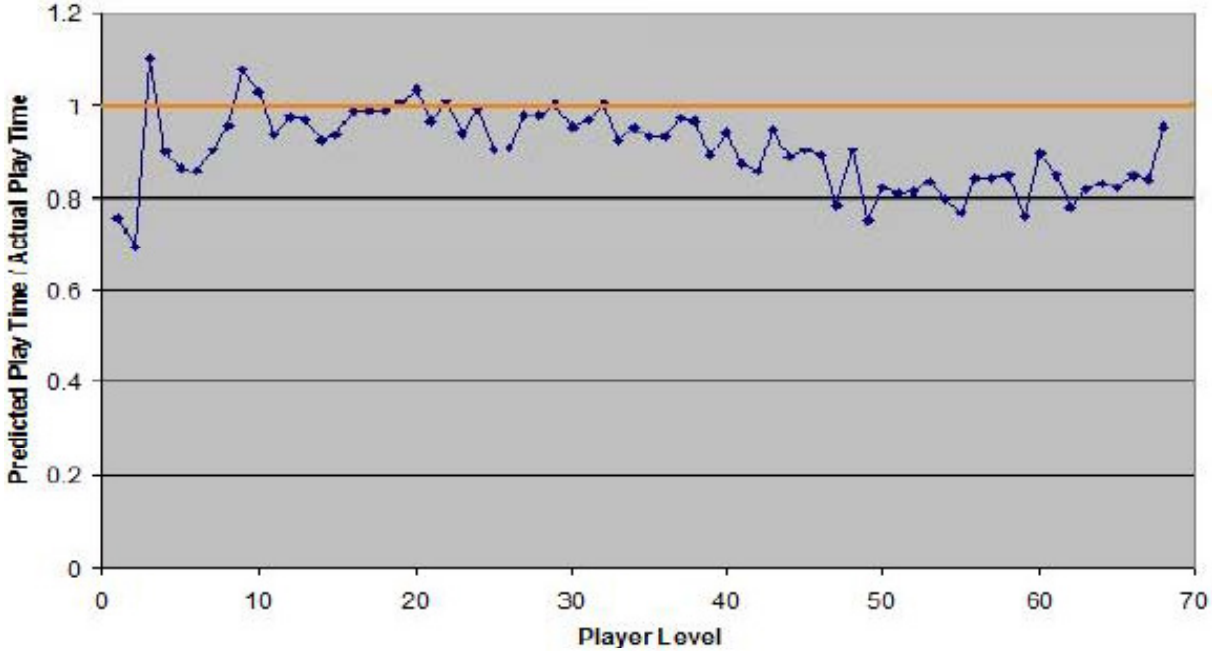


Fig. 5. Ratio of Predicted Play Time and Actual Play Time by Player Level

- We compare the predicted play time against the actual play time. We take the ratio between the two and observe at each player level what is the margin of prediction error is.
- Figure 5 shows that for Levels 2 and 3, our method underestimated the actual play time.
- Between Levels 4 and 48, the margin of prediction error stays well within 18% boundary.
- Beyond Level 48, the margin of error increases and players' performances become less predictable.
- For higher level players, our method tends to underestimate the actual play time.

Performance Metric 1 – Evaluation (more findings)

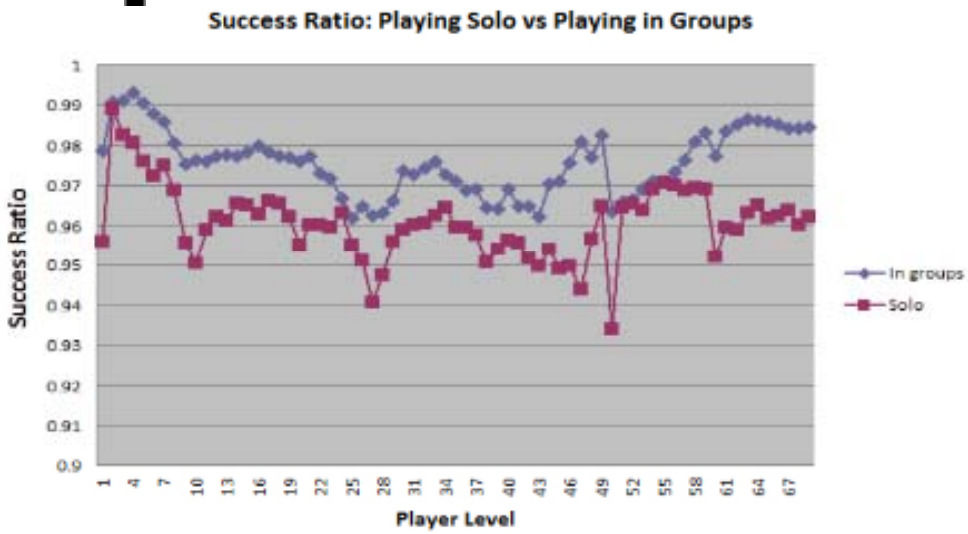


Fig. 3. Success Ratio - Playing Solo vs. Playing in Groups

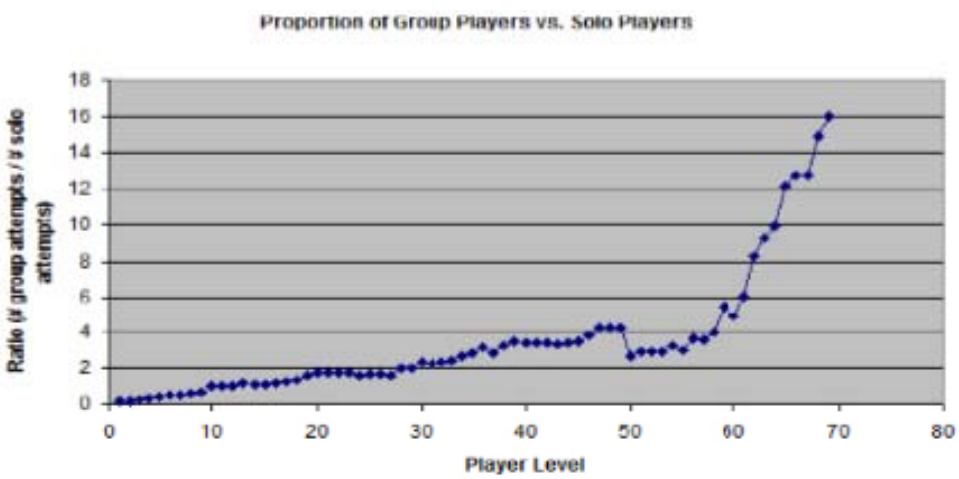


Fig. 4. Proportion of Group Players versus Solo Players

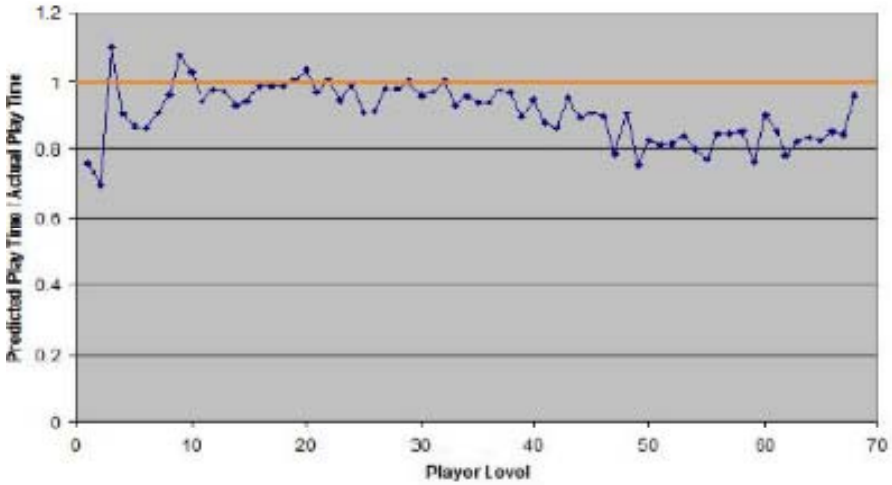


Fig. 5. Ratio of Predicted Play Time and Actual Play Time by Player Level

- As the player level increases, group formation becomes a more common occurrence. And playing in groups leads to higher success ratio at the individual player's level.

- There is a tradeoff between playing solo versus playing in groups. From timing perspective, playing solo allows a given player to advance faster than it would if he were to play in groups.
- From the perspective of successful task completion and success ratio, playing in groups serves as an advantage in that the chance of getting a given task done is higher for a given individual player in this setting.

Performance Metric 1 – Evaluation (distant past as a predictor of future performance)

Fig. 6. Predicted Play Time versus Actual Play Time (Distant past as a predictor of future performance)

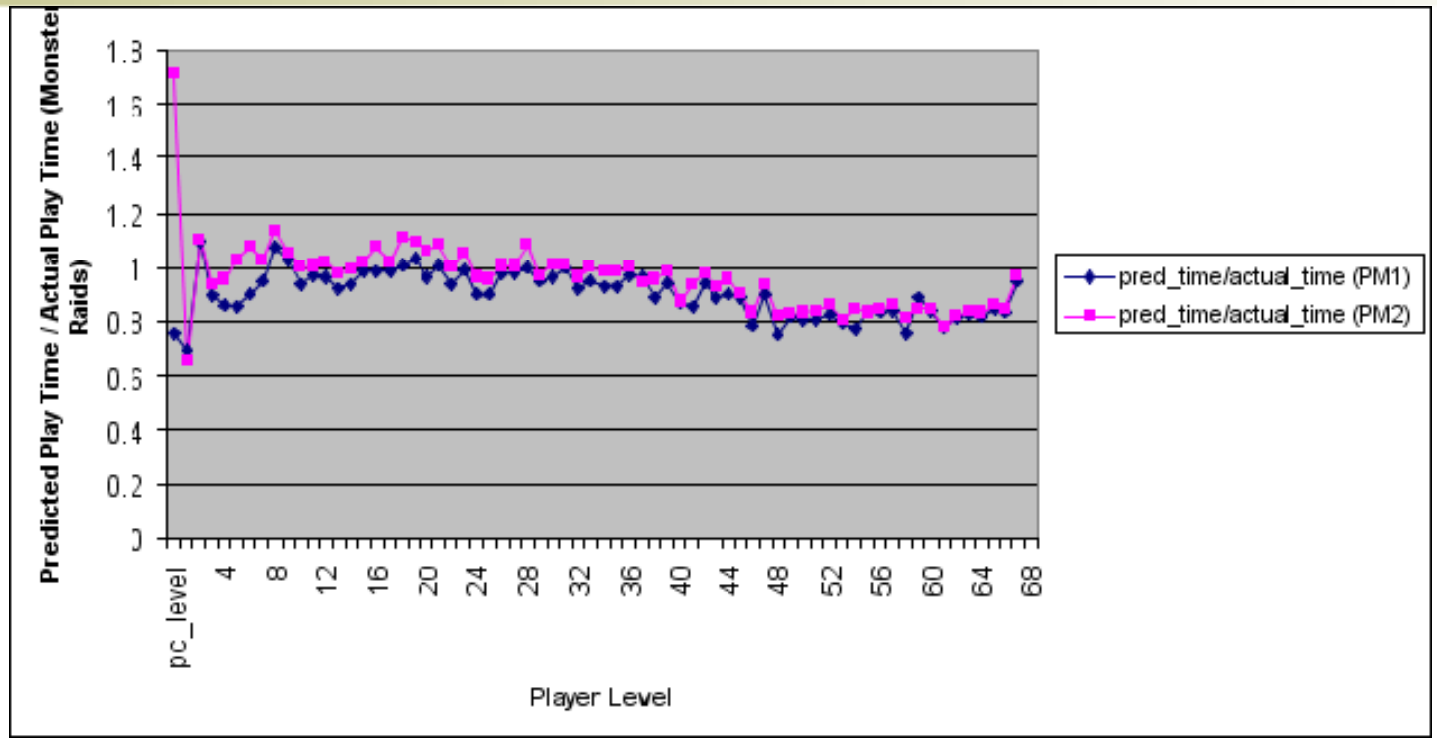
- In one use case, we used the immediate past performance data (from Level i + 1 to i + 2) as a predictor for future performance (from Level i+2 to i+3). **Pink plot.**
- In the other use case, we used the distant past performance data (from Level i to i + 1) as a predictor for future performance (from Level i + 2 to i + 3). **Blue plot.**

Performance Metric 1 – Evaluation (distant past as a predictor of future performance)

Fig. 6. Predicted Play Time versus Actual Play Time (Distant past as a predictor of future performance)

- Figure 6 shows that the margin of prediction error is larger when we use the distant past performance data as predictor for future performance. It is evident from the results that the quality of performance data as a predictor for future performance decays with an increasing distance on the player level scale.
- Our finding indicates that in order to incorporate more distant performance data into the proposed Performance Metrics method, some sort of a weight assignment or decay function must be applied in such a way that the most weight is given to the most immediate past.

Performance Metric 2 – Evaluation



- We conducted a similar experiment to evaluate Performance Metric 2.
- In this experiment, we used only monster kill tasks for analysis because these tasks are readily amenable to analysis in terms of success and failure. Failure in this context would mean the death in the game, while failure cannot be readily described for most other tasks.
- The results were not significantly different from the Performance Metric 1 evaluation results.
- In monster kills, taking quality into consideration for performance metric does not lead to better predictions.

Conclusion

- EverQuest II's existing ding points-based point scaling system is in general well in accordance with the actual player performance observed in the game's historical performance data.
  - The level of granularity that the ***performance data*** offers can potentially lead to fine tuning of the existing point scaling system
- Performance Metric 1 is suitable for predicting individual players' future performances in absence of impact of group formation.
- For certain type of task (i.e. monster kills), the quality aspect of individual performance plays an insignificant role in predicting player's future performance.

Future Directions

- More thorough and comprehensive studies on different types of group formations (homogeneous, heterogeneous, social interactions amongst the group members, etc.) and their impact on individual players' performances.
- Define quality in all types of tasks in the game for the purposes of devising more generalizable individual player performance metrics.
- Investigating individual learning trajectory over a larger time span (i.e. decay/weight model for leveraging distant performance data)
- Developing group performance metrics.

Thank you.

Q & A

Send questions to kjshim@cs.umn.edu