CALIFORNIA STATE UNIVERSITY, SACRAMENTO

PRESENTATION
on
RESEARCH PAPER AND PROJECT PROPOSAL

Presented by

DEVESH BINJOLA
NICHOLAS HANKS

Under the Guidance of

Dr. Meiliu Lu
Dept of Engineering And Computer Science

A RESEARCH PAPER ON

Mining Game Statistics from Web Services: A World of Warcraft
Armory case study

Chris Lewis - PhD, Computer Science
Noah Wardrip-Fruin - Professor of Computer Science

University of California, Santa Cruz

ABSTRACT

Collecting quantitative data sets from publicly available "web services" (in this case, the *WoW Armory* website) and mining the data for qualitative study

This study analyzed the following:
- (1) the ability of the constructed classifier to accurately predict character class based on equipment
- (2) the correlation between the time it takes players to reach the maximum level (80) and their chosen character class
- (3) the correlation between the number of deaths of a character and its class
- (4) the most popular items equipped among the sample

Process and Results

- To gather data, develop and implement a web crawler (WoWSpyder) to crawl *WoW Armory* website and download character data into database
- Filter out characters with errors (detected through "impossible" stats, such as negative deaths)
- Total sample size: 136,047 characters from the US and European servers

Process and Results

- **Naive Bayes classifier, 10-fold cross validation - 94.70% accuracy**

Confusion Matrix

	Druid	Warlock	Mage	Rogue	Paladin	Shaman	Warrior	Priest	DK	Hunter	Precision
P: Druid	**13286**	19	14	25	23	208	21	8	30	32	97.22%
P: Warlock	25	**8859**	980	5	20	14	6	831	4	4	82.42%
P: Mage	28	1407	**11153**	3	9	14	4	948	3	6	82.16%
P: Rogue	61	2	2	**11969**	6	9	7	0	0	82	98.61%
P: Paladin	0	0	2	1	**16882**	11	262	2	25	2	98.23%
P: Shaman	129	0	0	6	36	**10454**	3	0	3	11	98.23%
P: Warrior	2	0	0	5	384	10	**12301**	0	16	12	96.63%
P: Priest	28	311	265	0	5	7	1	**10239**	0	0	94.32%
P: DK	0	0	0	0	144	2	70	0	**19561**	0	98.91%
P: Hunter	124	1	0	146	15	277	25	1	1	**14138**	95.99%
Recall	97.10%	83.58%	89.83%	98.43%	96.34%	94.98%	96.86%	85.12%	99.58%	98.96%	

Purpose: discover deviations from designed "norms" in character equipment (class-targeted equipment)
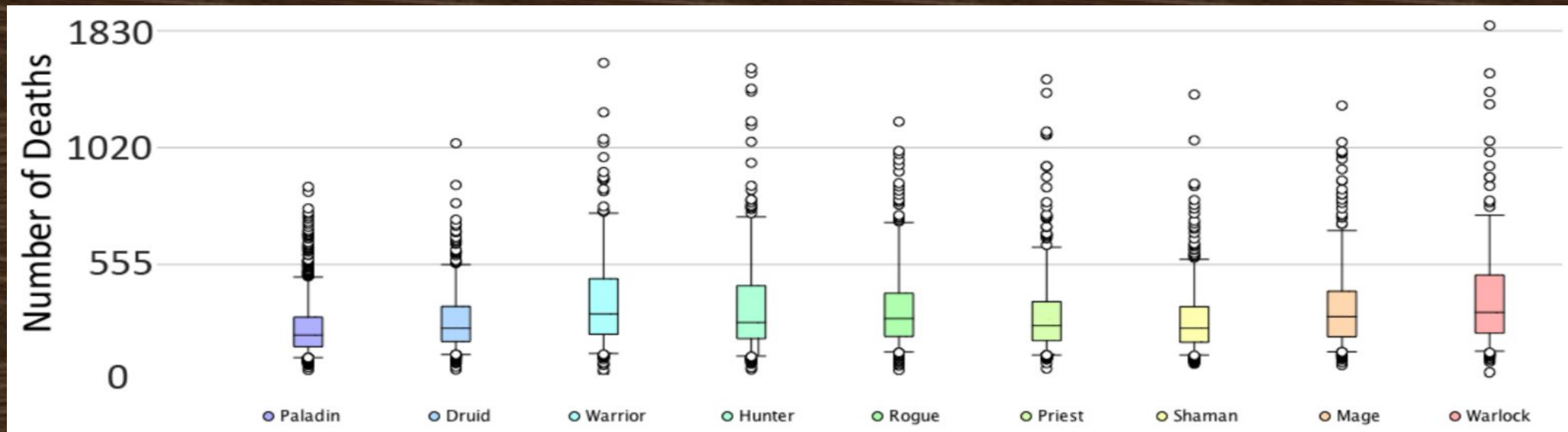
Process and Results

- Calculate time to reach maximum level (80) from level 10 for each character
- Construct box plot - Time (days) to Level 80 vs. Class
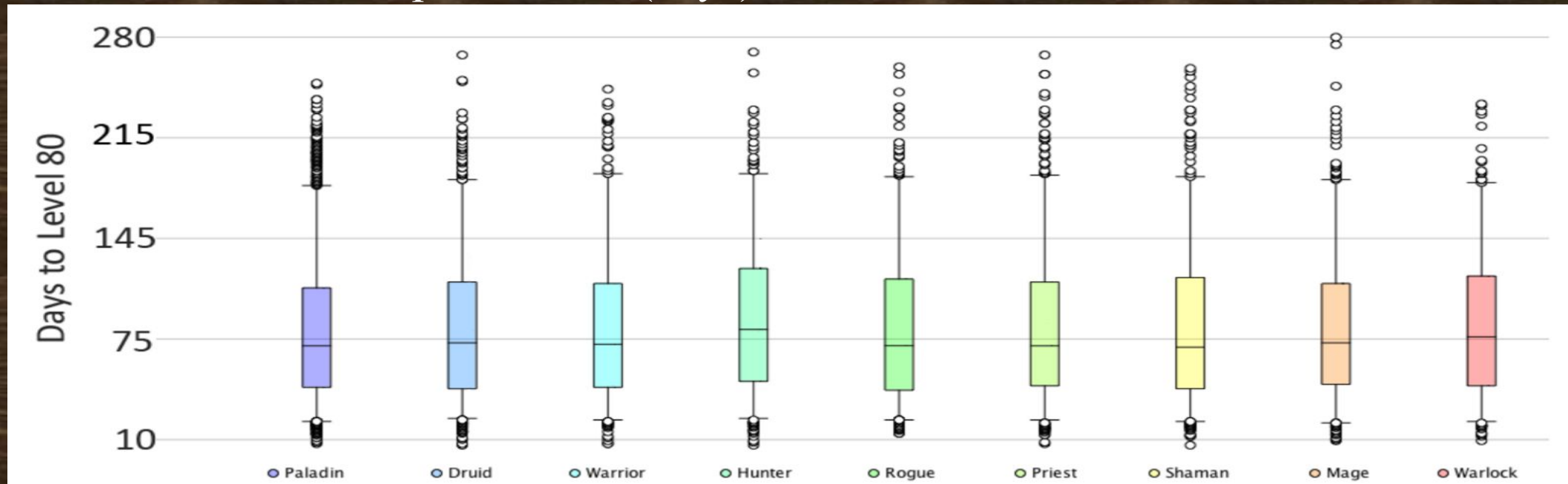
Analysis suggests there is no significant advantage/disadvantage among character classes

Process and Results

- Construct box plot - Number of Deaths vs. Class

Analysis suggests there is imbalance among character classes

Process and Results

- Construct list of most popular items (top 5 shown)

Item name	Number wearing
Guild Tabard	19849
Mirror of Truth	14219
Sigil of the Dark Rider	11574
Sundial of the Exiled	7505
Tabard of the Kirin Tor	7163

- Graph Item Popularity vs. Number Wearing
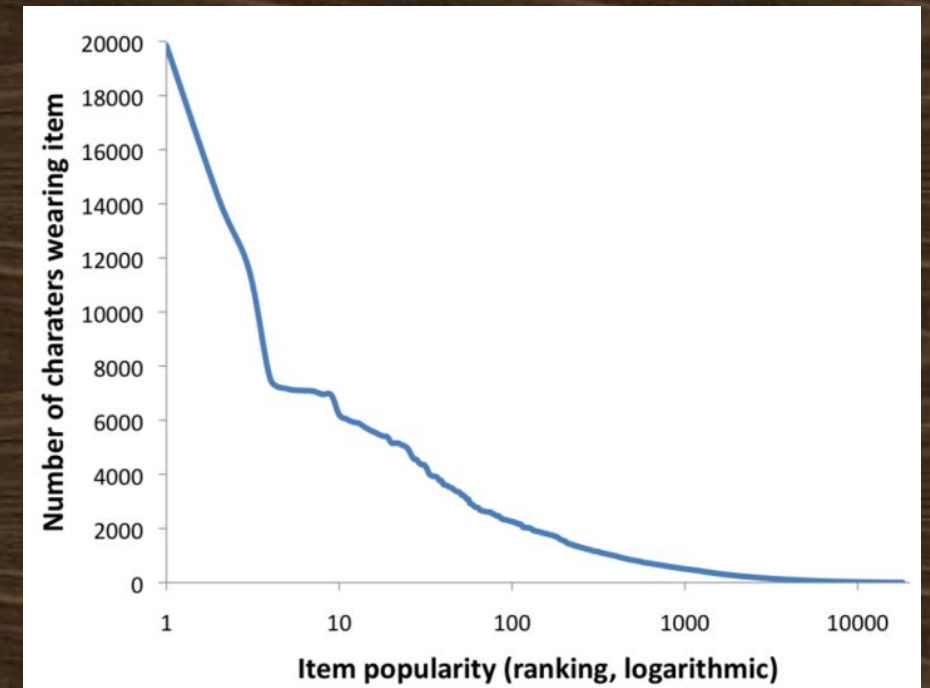
Analysis suggests there is a trade-off between item power and rarity

Limitations

- ▪ Web service query processing limitations (crawler restrictions)
  - – Workaround: multiple threads, built-in delays and error handling
- ▪ Financial limitations ($ no funding $)
- ▪ Technological limitations (1 machine collecting data)
- ▪ Data limitations (limited sample size, breadth of character data)
  - – Trimmed data records, focus on pertinent attributes

OPINION ON RESEARCH PAPER

A good study on how to gather and mine video game data from widely available web services. Video game player data can be difficult to obtain for casual researchers and those without industry connections. Given adequate resources, this study could be scaled to handle a larger volume of data, as well as more investment into discovering other problems that could be solved using data mining techniques.

DATA MINING PROJECT PROPOSAL

"STOCK MARKET PREDICTION USING DATA MINING"

PROJECT TEAM :

DEVESH BINJOLA
NICHOLAS HANKS

OBJECTIVE

The goal of this data mining project is to predict the behavior of the stock market and to trace out the trend such as high price, low price, volume etc. To understand these large data sets we make use of clustering algorithms.

    The algorithm used here is K-means. This project's aim is to find the best cluster since cluster points change whenever the application is executed. This is done by using the multiple random normalizations where it chooses the best initial clusters.

SCHEDULE

Total working of data mining may take up to a week. The tentative schedule would be like as follows.

1. Stock data collection – 2 days
2. Implementing the algorithm – 2 days
3. Analysis – 2 days

ALGORITHM USED

The algorithm used here is K-means. This project's aim is to find the best cluster since cluster points change whenever the application is executed. This is done by using the multiple random normalizations where it chooses the best initial clusters.

DATA WAREHOUSING Project Proposal

"DATAMART FOR GOLF DATABASE "

PROJECT TEAM :

DEVESH BINJOLA
NICHOLAS HANKS

OBJECTIVE

Main objective of this data warehousing project is to implement data mart which can help in searching the golf history from its beginning to till date(men's golf history).

This data mart can help to provide answers for the queries based on the database which are derived from the golf dataset:

The queries we are planning to implementing are:

1.Winner of majors by year

2.Count of majors by player name

3.Count of majors by major name

4.Winner of majors by major name

SCHEDULE

Total working of data ware housing may take upto 2 weeks. The tentative schedule would be like as follows.

1.golf data collection – 5 days

2.Data cleaning and preprocessing – 4 days

3.Implementation using PHP, MySQL – 5 days

REFERENCES

- WIKIPEDIA
- http://www.databasegolf.com/