Personal Project, Progress Report

Artem Vasenin

February 2, 2017

Email av429@cam.ac.uk

Project title Predicting arbitrary events in competitive computer team games

Supervisor Yingzhen Li (yl494)

Director of Studies Lawrence Paulson (lp15)

Overseers Hatice Gunes and Robert Watson

Algorithm outline

The pipeline of the final algorithm is as follows:

1. Collect and clean data

2. Get initial estimate of player skills of each variable by rating them individually with existing algorithms such as TrueSkill.

3. Train regressors and classifiers for each variable using initial estimates.

4. Update skill estimates using trained regressors and classifiers.

5. Train new regressors and classifiers using updated values.

6. Repeat previous two steps until peak accuracy is reached.

Work Completed

- All required data has been collected and cleaned.

- Different regression and classification methods have been tried, viz. neural networks, support vector machines, gaussian processes, linear, logistic, polynomial and ridge regression as well as naive bayes and random forests.

- General outline of the algorithm finished.

- Step one to three of the final algorithm completed.

- Biggest success currently was in predicting GPM and XPM variables achieving R^2 values of 0.43 and 0.35 respectively.

Schedule update

Project is about two weeks behind schedule, this can be remedied by changing the plan to following (Instead of improving accuracy in section 8, I will catch up):

7. **27 January - 9 February**

    Finish work on error estimation.

8. **10 - 23 February**

    Write code to update player skill statistic using error of predicted values, rather than using TrueSkill.

9. **24 February - 9 March**

    Optimising the code and correcting bugs. Learning how exactly to formally evaluate machine learning algorithms.

    Deliverables: Final version of the algorithm. Script that will be used for the evaluation of the algorithm.

10. **10 - 23 March**

    Aggregating completed work, summarising results and preparing an outline for the dissertation. Beginning the writing of the dissertation. Evaluating the final algorithm.

    Deliverables: Evaluation results for the final algorithm

11. **24 March - 6 April**

    Writing the dissertation.

12. **7 - 21 April1**

    Cleaning up the dissertation, making it presentable. Making sure it's clear and checking for errors. Complete first draft of the dissertation and ask the supervisor to check it.

    Deliverables: First draft of the dissertation.

13. **22 April - 4 May**

    Work with supervisor on correcting errors and improving the text.

    Deliverables: Finalised content of the dissertation.

14. **5 - 19 May**

    By this time the content should be finalised and cosmetic/presentation improvement should be made. Submit the dissertation before the deadline.

    Deliverables: Submission of the dissertation

Unexpected Difficulties

Some of the variables could not be predicted since results are missing from a lot of points in the dataset.

I believe that there may be a problem with achieving the success criterion of the original proposal, i.e. that mean squared error of predictions should be half the standard deviation in the dataset. That is the case since there appears to be not enough information in the type of data collected to predict many variables accurately.