DOTA2
GAME
PREDICTION

by
Yukai

School of Information Technology and Electrical Engineering,
The University of Queensland.

Submitted for the degree of
Bachelor of Engineering
in the field of ...

MONTH & YEAR.

ADDRESS LINE 1
ADDRESS LINE 2
Tel. (07) nnnn nnnn

October 19, 2017

Prof Paul Strooper
Head of School
School of Information Technology and Electrical Engineering
The University of Queensland
St Lucia, Q 4072

Dear Professor Strooper,

In accordance with the requirements of the degree of Bachelor of Engineering in the division of Electrical Engineering, Electrical and Biomedical Engineering, Electrical and Computer Engineering, Software Engineering, Mechatronic Engineering, I present the following thesis entitled "...". This work was performed [in partnership with Mr/Ms ... and] under the supervision of Mr/Ms/Dr/A/Prof./Prof. ....

I declare that the work submitted in this thesis is my own, except as acknowledged in the text and footnotes, and has not been previously submitted for a degree at The University of Queensland or any other institution.

Yours sincerely,

Author's Signature

AUTHOR'S NAME.

To . . .

Acknowledgments

Acknowledge your supervisor, preferably with a few short and specific statements about his/her contribution to the content and direction of the project. If you collaborated with another student, acknowledge your partner's contribution, including any parts of the thesis of which s/he was the principal author or co-author; this information can be duplicated in footnotes to the chapters or sections to which your partner has contributed. Briefly describe any assistance that you received from technical or administrative staff. Support of family and friends may also be acknowledged, but avoid sentimentality—or hide it in the dedication.

Abstract

This document is a skeleton thesis for 4th-year students. The printable versions (`skel.dvi, skel.ps, skel.pdf`) show the structure of a typical thesis with some notes on the content and purpose of each part. The notes are meant to be informative but not necessarily illustrative; for example, this paragraph is not really an abstract, because it contains information not found elsewhere in the document. The $\LaTeX\,2_\varepsilon$ source file (`skel.tex`) contains some non-printing comments giving additional information for students who wish to typeset their theses in $\LaTeX$. You can download the source, edit out the unwanted material, insert your own frontmatter and bibliographic entries, and in-line or `\include{}` your own chapter files. Of course the content of a particular thesis will influence the form to a large extent. Hence this document should not be seen as an attempt to force every thesis into the same mold. If in doubt about the structure of your thesis, seek advice from your supervisor.

Contents

Appendices 9

List of Figures

List of Tables

Chapter 1

Introduction

Dota2 is one of the most popular online game. It is also the originator of multiplayer online battle arena(MOBA) game. Professional Dota2 tournaments often offer prizes in millions of US dollars in total. There are over 6 millions of players who have played at least one match in February 2017. The popularity also brought a huge amount of data on players and matches. Valve, the developer of Dota2 produces API(Application Programming Interface) for people to access these data.

In a regular game of Dota, there are ten players separated to two sides. Each player can pick one of 113 heroes in the match. Players often play in different roles by convention. Each hero has different attribute and unique abilities which decide what roles it is capable of. When players pick heroes they need to consider the combination of roles and heroes, some powerful hero combos and countering enemies' pick. This is not easy because the game is extremely variable. For example: one hero can often be played as different roles; a certain item may increase a hero's strength or counter an enemy hero dramatically; the relationship between heroes: synergy or counter varies with time.

An experienced player can often predict the trend or even the result of a match by the composition of both teams — so called "draft". There are certain patterns between the "draft" and the result of the match, however, the total amount of the draft of a single team is ${}^{113}C_5 = 140364532$, and the combination of two teams is greater than 1.564×10^{16}. The complexity is huge, and machine learning is suitable for solving such a problem.

Machine learning learns patterns from a sample to solve problems. It has a variety of uses include credit applications, fraud detection, medical diagnosis, spam filters and many other applications[1]. There are different models for different types of problems. In this case, we have a classification problem. Logistic Regression and Naive Bayes are proposed to solve this problem due to their simplicity and efficiency.

Chapter 2

Literature review / prior art

Dota2 has drawn a fair amount of attention of machine learning researchers in the last few years due to its complexity and popularity.

In 2013 Conley and Perry presented a hero recommendation engine depends on the opposite draft[2]. They performed Logistic Regression and K-Nearest Neighbors on the draft information of a certain amount of matches. They achieved 69.8% accuracy with Logistic Regression on 18,000 matches, however, it couldn't learn from the relationships between heroes. To attempt solving that issue they used K-Nearest Neighbors with custom weights for neighbors with 2-fold cross-validation on 20,000 matches. Their best accuracy with 4 dimensions got 67.43% accuracy. The recommendation engine was built on these approaches. Their work was the first to use the draft information to predict Dota2 games result, but a linear model like Logistic Regression or simple model like K-Nearest Neighbors can not capture the interaction between heroes.

In 2014 Agarwala and Pearce studied how win rate depends on the draft by performing Logistic Regression with the draft directly, second order polynomial of PCA scores and PCA sorted by the first PCA component respectively on 1500 matches [3]. They selected some features to describe a hero's role by normalizing the average hero statistics. Then PCA was run on the average statistics of professional play. The result indicated that the PCA with draft models failed to match the accuracy of the pure draft model. The authors hypothesized that average hero stats are not so informative since one hero can play different roles. Their PCA score was based on professional play but they trained on public matches. Professional matches and public matches are very different and shouldn't be mixed, which should be another reason.

In 2015 Song et al.[4]. tried to use Logistic Regression to predict the winning side of Dota2 based on the draft. After discovering that only draft information is not enough to train a precise model, they attempted to add features based on powerful 2-hero combos using stepwise regression and the result was slightly better. They mentioned that the information of player should increase the model's accuracy in the conclusion.

In the same year Kinkade et al.[5] presented two classifiers to predicts the winning team of Dota2. One uses post game data, which performed perfectly but does not hold any real use. The other one uses pre-game draft information. The authors used Logistic Regression with several manually inserted features such as matchup, synergy and countering between heroes and they used Random Forest with only draft information. Logistic Regression got only 56% accuracy with draft information itself and the features increases the model's quality to up to 73%. For random forest, the authors found that overfitting was a large issue. They got 99% accuracy on the training set while they only got 67% accuracy on the testing set after several parameters verified. Logistic Regression performed significantly better. The authors hypothesized that the players' skill contributes almost linearly to the game so a linear model would fit better.

In 2016 Yang et al.[6] predict the winning team of a Dota2 match using prior and real-time data. The authors constructed a feature set covers more aspects including hero, player and hero-player combination as prior features and other real-time match stats as real-time features. They used Logistic Regression, neural network and the proposed Attribute Sequence Model and their combinations on prior features, and both prior and real-time features. For the prior game prediction, the result shows that using all features achieves the highest prediction accuracy 71.49%. For the real-time prediction, the results show that prior features' impact on the prediction drops over time while the real-time's increases. The idea of using the combination of information of hero and player is impressive.

In 2017 Neklyudov and Kirill[7] introduced Factorization Machines and Gradient Boosted Decision Trees for the prediction of Dota2 match with comparison to the models previously used by other researchers including Logistic Regression and Naive Bayes. The authors collected over 5 million matches from different game modes and skill levels. Then hero drafts were used as input of algorithms. The result indicated that Factorization Machines and Gradient Boosted Decision Trees performs better than other machine learning methods with the accuracy of 70.6% and 70.1% respectively in normal skill level. The authors also compared the performance of Gradient Boosted Decision Trees with and without role information and it performed constantly. The authors hypothesized that the algorithm somehow learned the features of heroes' roles so the additional information of roles did not impact much.

Chapter 3

Theory

3.1 Word2Vec

There are two ways of implementing Word2vec[8]: skip-gram model and negative sampling. Although it was using negative sampling in our implementation, we will explain how it works with skip-gram model, since the theory behind negative sampling is essentially skip-gram model.

We are given a corpus of words w and their contexts c within a window size. Given a corpus $Text$, the goal is to derive the parameters θ to maximize the corpus probability[9]:

$$\arg\max_{\theta} \prod_{(w,c)\in D} p(c|w,\theta)$$

In our case, a word w is a hero, its context c is its teammates. D is the set of all two-hero pairs in the same team. The corpus are all teams in our trainning data. The window size is 4 since we are certain that a hero interacts with all 4 of its teammates.

Chapter 4

Methodology, procedure, design, etc.

This may be one chapter or several. Again, titles should be more informative than the above.

You will almost certainly need diagrams to clarify your meaning. The LaTeX 2ε `graphics` package allows the inclusion of PostScript graphics, as in Fig. 4.1. The inclusion of LaTeX `picture` graphics, as in Fig. 4.2, requires no auxiliary packages and allows the mathematical formatting features of LaTeX to be used in diagrams; but the `picture` files, unlike PostScript files, usually require manual editing.

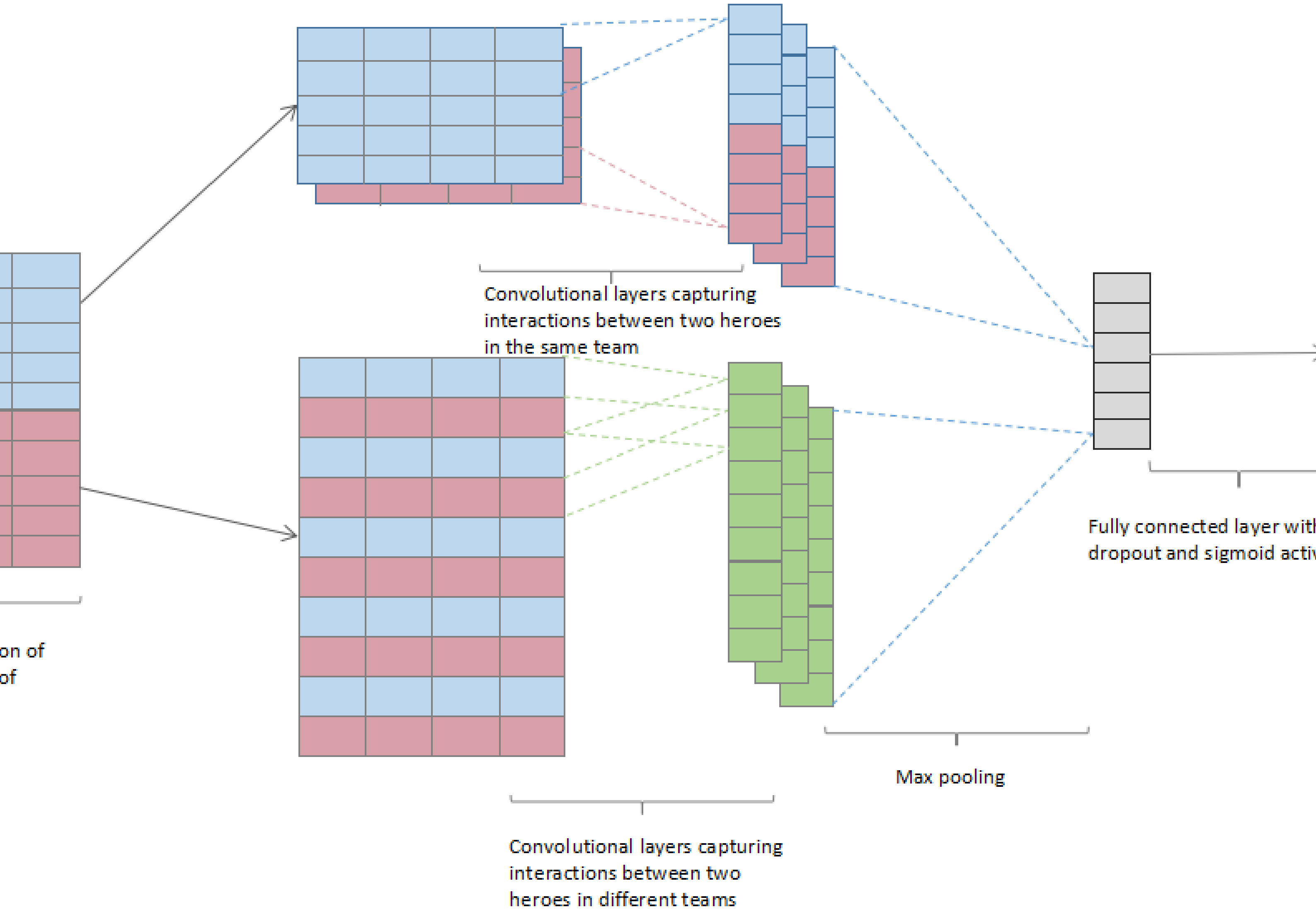
Convolutional layers capturing interactions between two heroes in the same team
on of
of
Fully connected layer wit
dropout and sigmoid acti
Max pooling
Convolutional layers capturing interactions between two heroes in different teams

Figure 4.1: Experimental two-way active crossover (op-amp version)

Figure 4.2: Modeling a discrete-time LTI system using z-transforms

Chapter 5

Results and discussion . . .

. . . or perhaps the discussion should be a separate chapter.

In any case, you will probably need to include tabulated results. Table 5.1 illustrates the use of various LaTeX environments to include a computer printout (plain text file) in a document. The `verbatim` environment, which encloses the formatted text, is also useful for program listings.

Table 5.1: *Fraction of air volume involved in heat exchange for second mode (right column) vs. filling factor (left column). The plain-text headings represent f, m, μ_2 and f_2.*

```
 f(%)     m         mu2     f2(%)

 0.016   80.00    0.05400   4.874
 0.031   56.57    0.07732   5.438
 0.062   40.00    0.11103   6.125
 0.125   28.28    0.16001   6.970
 0.250   20.00    0.23175   8.020
 0.500   14.14    0.33799   9.329
 1.000   10.00    0.49789  10.967
 2.000    7.07    0.74444  13.008
 4.000    5.00    1.13919  15.525
 8.000    3.54    1.81095  18.568

19.237    2.28    3.61958  23.174
37.180    1.64    7.28635  27.094
57.392    1.32   14.63631  29.813
74.316    1.16   29.35160  31.453
85.734    1.08   58.79364  32.360
```

Chapter 6

Conclusions

6.1 Summary and conclusions

6.2 Possible future work

Appendix A

Dummy appendix

Appendices are useful for supplying necessary details or explanations which do not seem to fit into the main text, perhaps because they are too long and would distract the reader from the central argument. Appendices are also used for program listings.

Notice that appendices are “numbered” with capital letters, not numerals. When the `\appendix` command in LaTeX [**?**, p. 175] is used with the `book` document class, it causes subsequent chapters to be treated as appendices.

Appendix B

Program listings

B.1 First program

Some initial explanatory notes may precede the listing.

B.2 Second program

B.3 Etc.

Appendix C

Companion disk

If you wish to make some computer files available to your examiners, you can list and describe the files here. The files can be supplied on a disk and inserted in a pocket fixed to the inside back cover.

The disk will not be needed if you can specify a URL from which the files can be downloaded.

Bibliography

[1] E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.

[2] K. Conley and D. Perry, "How does he saw me? a recommendation engine for picking heroes in dota 2," *Np, nd Web*, vol. 7, 2013.

[3] A. Agarwala and M. Pearce, "Learning dota 2 team compositions," tech. rep., Stanford University, Tech. Rep., 2014.

[4] K. Song, T. Zhang, and C. Ma, "Predicting the winning side of dota2," tech. rep., Stanford University, Tech. Rep., 2015.

[5] N. Kinkade, L. Jolla, and K. Lim, "Dota 2 win prediction," tech. rep., University of California San Diego, Tech. Rep., 2015.

[6] Y. Yang, T. Qin, and Y.-H. Lei, "Real-time esports match result prediction," *arXiv preprint arXiv:1701.03162*, 2016.

[7] K. Neklyudov, "Performance of machine learning algorithms in predicting game outcome from drafts in dota 2," in *Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016, Yekaterinburg, Russia, April 7-9, 2016, Revised Selected Papers*, vol. 661. Springer, 2017, p. 26.

[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[9] Y. Goldberg and O. Levy, "word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method," *arXiv preprint arXiv:1402.3722*, 2014.

[10] [Online]. Available: https://www.kaggle.com/devinanzelmo/dota-2-matches

[11] [Online]. Available: https://www.kaggle.com

[12] [Online]. Available: http://comprehensivepainmanagementcenter.com/correct-sitting-posture/

[13] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.