Data Stream Clustering Applied to Yugioh Data

Vincent Allen

5/8/2020

cs6405: Clustering Algorithms

About the Author:

After graduating from the Missouri University of Science and Technology with an Electrical Engineering B.S degree in Dec 2014, Vincent joined Boeing as a Software Engineer II. He has since started a Computer Science M.S. degree and plans to graduate in Dec 2020. During his into to AI course, Vince’s [AI](https://github.com/vtad4f/chess-ai) (Volcanic Counter) placed 3rd in the chess tournament that semester. Later on, in his free time he created a [UI](https://github.com/vtad4f/chess-ui) in python to go with it. Growing up Vince enjoyed playing the trading card game Yugioh.

Table of Contents

[Table of Contents 2](#_Toc39752604)

[Executive Summary 3](#_Toc39752605)

[Introduction 3](#_Toc39752606)

[Project Specifications 3](#_Toc39752607)

[Detailed Design 3](#_Toc39752608)

[Experimental Results 3](#_Toc39752609)

[References 4](#_Toc39752610)

[Appendix A 4](#_Toc39752611)

[Appendix B 4](#_Toc39752612)

# Executive Summary

Before proceeding, we’ll give a brief but relevant introduction to Yugioh. It is a trading card game with thousands of cards and variety of card types. Imagine two children saying back and forth “My alien shoots your dinosaur with a lazer gun”, then “Well my dinosaur uses an anti lazer shield”, then “My alien eats shields” and you have a good idea of what a game of Yugioh is like. There is enough variety in the card effects to make analysis of them difficult. There will always be outliers.

The goal of this project is to use stream clustering to predict the level of a Yugioh monster card based on its attack, defense, and card text. Most monster types have a level/rank between 0 and 12. There isn’t a rule that says higher level monsters should be stronger than lower level monsters (i.e. have higher atk and def or better effects), but that’s the general trend. In most cases low level monsters are sacrificed to summon high level monsters, so this makes sense. In any case, the interesting aspect of the clustering should be translating the card text into numerical attributes.

With stream clustering we are able to make a prediction as we process every single point of data, instead of having to wait for a training phase to finish. Thus, we are able to watching metrics update dynamically as the algorithm walks the data.

‘Normal’ monster text will be omitted as it serves no functional purpose. Text that actually gives the monster some effect on the game will be preprocessed into numeric values to be used for distance calculations.

# Introduction

*Here you will need to frame the project both within the particular subfield and within the larger framework of engineering as a whole. Assume your reader is a technical professional with little knowledge of this course’s subject but has general technical knowledge. Students often struggle here wondering exactly how much detail or background information to include. Try searching ieeexplore.org for professional research journal articles that interest you. Read the introduction and literature summary portions of the papers. This will give you some idea of what is needed in a technical introduction.*

*This section should be 2-5 pages in length and should be at a high level. Leave the full technical details for later sections. This and the subsequent sections can include figures.*

# Project Specifications

*Just a very brief description of what you are trying to accomplish.*

*You will find as you get into the writing of this that your ability to understand and explain the project increases. This makes your work more accessible and, since it can be communicated, more likely to actually have a technological impact.*

# Detailed Design

*This section will form the bulk of the report. Here you should include design decisions and tradeoffs as well as any detailed technical drawings such as circuit diagrams, flow charts, etc. If these are particularly large they may be placed in an appendix, but should be referenced in this section.*

*It is here that you really get into the details of why your project is designed the way it is. Tradeoffs are made in a number of areas and a good way to organize this section is to figure out what the most important tradeoffs are and explain each of them with a few paragraphs.*

*This is a section that will evolve as the project nears completion, but its writing could be started at the beginning of the semester. I’ll be happy to give feedback on whatever you are able to produce for this section during the semester, while recognizing that some elements will be subject to change or impossible to write up until the final project is completed.*

# Experimental Results

*At some point you will have to determine whether your project works. This section should detail the design of the testing experiments, the results of the testing, and comments regarding whether the project does what it is supposed to do. Elements that are hard to test and aspects of the project that do not pass the tests should be highlighted. The design of the experiments is worth some space as well, as there are design tradeoffs and decisions to be discussed here as well as for the project itself. Any data should be included in tables or charts in this section or in an extra appendix (in addition to the one described below) if it is particularly cumbersome.*

# References

*Just like in a publication. Use IEEE style references.*

# Appendix A

*Give both code and pseudocode. The former should be plain text and the latter should be formatted the same as the rest of the report. You are highly encouraged, but not required, to share the code on GitHub. This will not affect your grade but it is usually good for your career. If you do share the code, please provide the link.*

# Appendix B

See additional file for the annotated bibliography.