Cluster Analysis Using N-gram Statistics for Daihinmin Programs and Performance Evaluations

Seiya Okubo, School of Management & Information, University of Shizuoka, Shizuoka, Japan
Takaaki Ayabe, Graduate School of Electro-Communications, University of Electro-Communications, Chofu, Japan
Tetsuro Nishino, Graduate School of Informatics and Engineering, University of Electro-Communications, Chofu, Japan

ABSTRACT

In this paper, the authors elucidate the characteristics of the computer game Daihinmin, a popular Japanese card game that uses imperfect information. They first propose a method to extract feature values using n-gram statistics and a cluster analysis method that employs feature values. By representing the program card hands as several symbols, and the order of hands as simplified symbol strings, they obtain data that is suitable for feature extraction. The authors then evaluate the effectiveness of the proposed method through computer experiments. In these experiments, they apply their method to ten programs that were used in the UEC Computer Daihinmin Convention. In addition, the authors evaluate the robustness of the proposed method and apply it to recent programs. Finally, they show that their proposed method can successfully cluster Daihinmin programs with high probability.

KEYWORDS
Daihinmin, Dendrogram, Imperfect Information Games, N-gram Statistics

1. INTRODUCTION

Many attempts have been made to enable a computer to play games. In some games, such as Go or Japanese chess, computer brains have reached the abilities of human professional players. These games are perfect information games, in which the players are given equal amounts of information.

In imperfect information games, on the other hand, information is partially hidden from the players. Studies conducted on imperfect information games, such as mahjong, have involved massive data analyses. Another imperfect information game is Daihinmin (or Daifugo), a Japanese card game that is similar to President, a Western card game. Computer Daihinmin is the practice of using computers to play Daihinmin. Since 2006, the University of Electro-Communications Computer Daihinmin
Convention (UECda) has taken place to facilitate this practice (Nishino & Okubo, 2009). At the convention, attendees can study algorithms used for imperfect information games, and researchers can propose algorithms of this type (Konuma, Honda, Hoki & Nishino, 2012; Suto & Shinohara, 2009; Suto, Narisawa & Shinohara, 2010). Outcomes from these studies are reflected in UECda participating programs, which are considerably improved each year.

Mainstream powerful player programs employ the Monte Carlo method or other randomized algorithms. When these programs are played, even the developers who wrote the programs cannot predict the next hand. Furthermore, owing to the high-speed matches that characterize the convention, as well as the absence of human professional players in the game, there are no established game-playing tactics or styles, such as those often observed in popular games.

The purpose of this study is to extract feature quantities of Daihinmin programs. We herein propose an extraction procedure using n-gram statistics and a cluster analysis method. We validate the proposed technique by applying it to UECda-featured programs.

2. PRELIMINARY

Various studies have been conducted on the similarity of programs, such as comparisons of source code and of feature quantity behavior and extraction (Kikuchi, Goto, Wakatsuki & Nishino, 2014; Nakamura, 1997; Yamada & Mizuno, 2014). In this study, we use n-gram statistics and Ward’s method.

2.1. N-gram Statistics

N-gram statistics comprise a linguistic model of the type and emergence ratio of \( n \) items of element strings appearing next to a sequence of words or letters. N-gram statistics are typically used to extract idioms and identify authors in the area of natural language processing. In addition, these statistics are widely applied to extensive areas of research. By utilizing them in game studies, we can extract procedures with different lengths and identify established procedures, such as behavioral choices unconsciously made by players, without placing constraints on the game.

In studies of perfect information games, n-gram statistics are used to automatically extract established patterns (Nakamura, 1997). For Daihinmin, for example, no clustering or other detailed analyses have been made to date.

2.2. Cluster Analysis

In this study, we use three types of distance concepts to compute distances. We then use Ward’s method for clustering. The computation methods for each distance concept are defined below.

**Manhattan distance:** Manhattan distance can be obtained by measuring the distance between two points along orthogonal coordinate axes. It is defined as Equation 1:

\[
 d_1(x, y) = \sum_{i=1}^{n} |x_i - y_i| 
\]

**Euclidean distance:** It is a distance that is applicable in Euclidean space. It is defined as Equation 2:

\[
 d_2(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} 
\]
IDS Using Reinforcement Learning Automata for Preserving Security in Cloud Environment
[www.igi-global.com/article/ids-using-reinforcement-learning-automata-for-preserving-security-in-cloud-environment/205594?camid=4v1a](www.igi-global.com/article/ids-using-reinforcement-learning-automata-for-preserving-security-in-cloud-environment/205594?camid=4v1a)

Model-Driven Automated Error Recovery in Cloud Computing
[www.igi-global.com/chapter/model-driven-automated-error-recovery/49157?camid=4v1a](www.igi-global.com/chapter/model-driven-automated-error-recovery/49157?camid=4v1a)
Attribute Decoration of Attack–Defense Trees
www.igi-global.com/article/attribute-decoration-attack-defense-trees/66406?camid=4v1a

Globalising Software Development in the Local Classroom
www.igi-global.com/chapter/globalising-software-development-local-classroom/29496?camid=4v1a