



Integration of ART2 neural network and genetic K-means algorithm for analyzing Web browsing paths in electronic commerce

R.J. Kuo*, J.L. Liao, C. Tu

Department of Industrial Engineering and Management, National Taipei University of Technology, No. 1, Section 3, Chung-Hsiao East Road, Taipei 106, Taiwan, ROC

Received 15 May 2003; received in revised form 5 February 2004; accepted 14 April 2004
Available online 2 June 2004

Abstract

Neural networks and genetic algorithms are useful for clustering analysis in data mining. Artificial neural networks (ANNs) and genetic algorithms (GAs) have been applied in many areas with very promising results. Thus, this study uses adaptive resonance theory 2 (ART2) neural network to determine an initial solution, and then applies genetic K-means algorithm (GKA) to find the final solution for analyzing Web browsing paths in electronic commerce (EC). The proposed method is compared with ART2 followed by K-means.

In order to verify the proposed method, data from a Monte Carlo Simulation are used. The simulation results show that the ART2+GKA is significantly better than the ART2+K-means, both for mean within cluster variations and misclassification rate. A real-world problem, a recommendation agent system for a Web PDA company, is investigated. In this system, the browsing paths are used for clustering in order to analyze the browsing preferences of customers. These results also show that, based on the mean within-cluster variations, ART2+GKA is much more effective.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Clustering analysis; Data mining; ART2; Genetic K-means algorithm; Recommendation agent system

1. Introduction

Electronic commerce (EC) has developed rapidly in recent years. Because entering the Internet is not difficult and creating customer interaction is easy on the Internet, how to create long-term customer relationships is a critical factor for successful EC. The Institute for Information Industry (III) showed that there were over six million Internet users by the end of

2000, and it has increased dramatically. This results in more requirements for the analysis of network loading and more complexity of Web site design.

To help users in browsing Web contents is an important factor for designing a Web site. Thus, browsing behavior becomes an important index of effectiveness of a site. By analyzing the frequency of Web page clicks and by understanding the regular browsing paths, the structure of Web site can be improved and more popular Web pages can be provided to the customers, thereby increasing EC sales.

* Corresponding author. Fax: +886-2-27712171x4541.
E-mail address: rjkuo@ntut.edu.tw (R.J. Kuo).

Therefore, this research proposes a novel clustering analysis technique for data mining. It is employed to analyze the browsing paths and behaviors of EC customers in order to improve the development of Web sites and customer satisfaction. Clustering analysis is a common tool in multivariate analysis and has been applied widely in many areas. The purpose of clustering analysis is to determine the objects in the same cluster with similar characteristics. Furthermore, it determines when there is a significant difference between two different clusters. The applications of clustering analysis include social science, genetics, biology, business and education.

In addition to statistical methods, artificial neural networks (ANNs) have also been widely applied in such areas. The unsupervised neural network, which is able to cluster objects by learning from training samples, is especially useful. Genetic algorithms (GAs) also have this capability. Kuo et al. [11] have reported that very good solutions can be provided by using self-organizing feature maps (SOM) of neural networks to determine the number of clusters and the starting points, and then employing the K-means method to find the final solution. They also showed that K-means can be replaced by GA in order to get better results [11]. [10] Thus, this current study proposes using a modified two-stage clustering method, adaptive resonance theory 2 (ART2) neural network [3,4], followed by genetic K-means algorithm (GKA). The results are compared with those from ART2 followed by K-means.

In order to evaluate the performance of the two clustering analysis methods, both simulation and real-world data are employed. Both of sets of results indicate that the proposed method is better than ART2 followed by K-means. Then, the proposed method is applied to create a recommendation agent system for a PDA Web company, which can dramatically increase customer satisfaction.

The remainder of this paper is organized as follows. Section 2 introduces some necessary background, including applications of neural networks and genetic algorithms for clustering analysis. The proposed method is presented in Section 3, and Section 4 illustrates the simulation results. Section 5 shows the model evaluation results for a PDA Web company, and concluding remarks are made in Section 6.

2. Background

This section introduces some general background of data mining. In addition, applications of neural networks and genetic algorithms for clustering are also presented.

2.1. Data mining

Data mining is a popular research topic in EC, primarily used to extract usable information from huge data sets to more quickly make effective decisions. As one aspect of Knowledge Discovery in Databases (KDD) [20], it is able to find relevant patterns from huge data sets and to analyze usable information to aid in making choices. The concept of data mining is not simply regression analysis or a database management system (DBMS). To be a form of regression analysis, it should have a basic hypothesis and be able to process huge data set, whereas DBMS cannot analyze data to obtain more advanced information. Only data mining techniques, which combine statistics, data warehouse and machine learning, can act as a valuable tool for commercial purposes. Previously proposed methods for data mining can be categorized as follows: (1) association rules, (2) decision tree, (3) cluster analysis, (4) induction method, (5) statistic, (6) online analysis process or OLAP, (7) neural networks and (8) genetic algorithms [14]. This study will focus on third part, cluster analysis. Thus, applications of ANNs and genetic algorithms are summarized in the following two subsections. This study intends to realize the relationship between customer's browsing paths and preferences.

2.2. The application of ANNs in clustering analysis

An artificial neural network (ANN) is a system which has been derived through models of neurophysiology. In general, it consists of a collection of simple nonlinear computing elements whose inputs and outputs are linked to form the network.

The learning algorithms of ANNs can be divided either: supervised and unsupervised [5]. In supervised learning, the network has its output compared with a known answer and receives feedback about any errors. This is sometimes called learning with a

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات