Expert Systems with Applications 37 (2010) 5621–5629

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A novel clustering algorithm based upon games on evolving network

Qiang Li [a,*], Zhuo Chen [b], Yan He [a], Jing-ping Jiang [a]

[a] College of Electrical Engineering, Zhejiang University, Hang Zhou, Zhejiang 310027, China
[b] Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Keywords:
Unsupervised learning
Data clustering
Evolutionary game theory
Evolutionarily stable strategy

ABSTRACT

This paper introduces a model based upon games on an evolving network, and develops three clustering algorithms according to it. In the clustering algorithms, data points for clustering are regarded as players who can make decisions in games. On the network describing relationships among data points, an edge-removing-and-rewiring (ERR) function is employed to explore in a neighborhood of a data point, which removes edges connecting to neighbors with small payoffs, and creates new edges to neighbors with larger payoffs. As such, the connections among data points vary over time. During the evolution of network, some strategies are spread in the network. As a consequence, clusters are formed automatically, in which data points with the same evolutionarily stable strategy are collected as a cluster, so the number of evolutionarily stable strategies indicates the number of clusters. Moreover, the experimental results have demonstrated that data points in datasets are clustered reasonably and efficiently, and the comparison with other algorithms also provides an indication of the effectiveness of the proposed algorithms.

1. Introduction

Cluster analysis is an important branch of Pattern Recognition, which is widely used in many fields such as pattern analysis, data mining, information retrieval and image segmentation. For the past thirty years, many excellent clustering algorithms have been presented, say, *K*-means (MacQueen, 1967), C4.5 (Quinlan, 1993), support vector clustering (SVC) (Ben-Hur, Horn, Siegelmann, & Vapnik, 2001), spectral clustering (Ng, Jordan, & Weiss, 2002), etc., in which the data points for clustering are fixed, and various functions are designed to find separating hyperplanes. In recent years, however, a significant change has been made. Some researchers thought about why not those data points could move by themselves, just like agents or something, and collect together automatically. Therefore, following their ideas, they created a few exciting algorithms (Cui, Gao, & Potok, 2006; Folino & Spezzano, 2002; Labroche, Monmarché, & Venturini, 2003; Rhouma & Frigui, 2001; van der Merwe & Engelbrecht, 2003), in which data points move in space according to certain simple local rules preset in advance.

Game theory came into being with the book named "Theory of Games and Economic Behavior" by John von Neumann and Oskar Morgenstern (Neumann & Morgenstern, 1944) in 1940. In this period, Cooperative Game was widely studied. Till 1950's, John Nash published two well-known papers to present the theory of non-cooperative game, in which he proposed the concept of Nash equilibrium, and proved the existence of equilibrium in a finite non-cooperative game (Nash, 1950, 1951). Although non-cooperative game was established on rigorous mathematics, it required that players in a game must be perfect rational or even hyper-rational. If this assumption could not hold, the Nash equilibrium might not be reached sometimes. On the other hand, evolutionary game theory (Smith, 1976) stems from the researches in biology which are to analyze the conflict and cooperation between animals or plants. It differs from classical game theory by focusing on the dynamics of strategy change more than the properties of strategy equilibria, and does not require perfect rational players. Besides, an important concept, evolutionarily stable strategy (Smith, 1976; Smith & Price, 1973), in evolutionary game theory was defined and introduced by Smith and Price in 1973, which was often used to explain the evolution of social behavior in animals.

To the best of our knowledge, the problem of data clustering has not been investigated based on evolutionary game theory. So, if data points in a dataset are considered as players in games, could clusters be formed automatically by playing games among them? This is the question that we attempt to answer. In our clustering algorithm, each player hopes to maximize his own payoff, so he constantly adjusts his strategies by observing neighbors' payoffs. In the course of strategies evolving, some strategies are spread in the network of players. Finally, some parts will be formed automatically in each of which the same strategy is used. According to different strategies played, data points in the dataset can be naturally collected as several different clusters. The remainder of this paper is organized as follows: Section 2

* Corresponding author.
E-mail addresses: anjuh@cqu.edu.cn (Q. Li), czcq@sjtu.edu.cn (Z. Chen), heyan@zju.edu.cn (Y. He), eejiang@zju.edu.cn (J.-p. Jiang).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.02.050

introduces some basic concepts and methods about the evolutionary game theory and evolutionary game on graph. In Section 3, the model based upon games on evolving network is proposed and described specifically. Section 4 gives three algorithms based on this model, and the algorithms are elaborated and analyzed in detail. Section 5 introduces those datasets used in the experiments briefly, and then demonstrates experimental results of the algorithms. Further, the relationship between the number of clusters and the number of nearest neighbors is discussed, and three edge-removing-and-rewiring (ERR) functions employed in the clustering algorithms are compared. The conclusion is given in Section 6.

2. Related work

Cooperation is commonly observed in genomes, cells, multi-cellular organisms, social insects, and human society, but Darwin's Theory of Evolution implies fierce competition for existence among selfish and unrelated individuals. In past decades, many efforts have been devoted to understanding the mechanisms behind the emergence and maintenance of cooperation in the context of evolutionary game theory.

Evolutionary game theory, which combines the traditional game theory with the idea of evolution, is based on the assumption of bounded rationality. On the contrary, in classical game theory players are supposed to be perfectly rational or hyper-rational, and always choose optimal strategies in complex environments. Finite information and cognitive limitations, however, often make rational decisions inaccessible. Besides, perfect rationality may cause the so-called backward induction paradox (Pettit & Sugden, 1989) in finitely repeated games. On the other hand, as the relaxation of perfect rationality in classical game theory, bounded rationality means people in games need only part rationality (Simon, 1996), which explains why in many cases people respond or play instinctively according to heuristic rules and social norms rather than adopting the strategies indicated by rational game theory (Szabó & Fath, 2007). So, various dynamic rules can be defined to characterize the boundedly rational behavior of players in evolutionary game theory.

Evolutionary stability is a central concept in evolutionary game theory. In biological situations the evolutionary stability provides a robust criterion for strategies against natural selection. Furthermore, it also means that any small group of individuals who tries some alternative strategies gets lower payoffs than those who stick to the original strategy (Weibull, 1995). Suppose that individuals in an infinite and homogenous population who play symmetric games with equal probability are randomly matched and all employ the same strategy A. Nevertheless, if a small group of mutants with population share $\epsilon \in (0,1)$ who plays some other strategy appear in the whole group of individuals, they will receive lower payoffs. Therefore, the strategy A is said to be evolutionary stable for any mutant strategy B, if and only if the inequality, $E(A,(1-\epsilon)A+\epsilon B) > E(B,(1-\epsilon)A+\epsilon B)$, holds, where the function $E(\cdot,\cdot)$ denotes the payoff for playing strategy A against strategy B (Broom, Cannings, & Vickers, 2000).

In addition, the cooperation mechanism and spatial-temporal dynamics related to it have long been investigated within the framework of evolutionary game theory based on the prisoner's dilemma (PD) game or snowdrift game which models interactions between a pair of players. In early days, the iterated PD game was widely studied, in which a player interacted with all other players. By round Robin interactions among players, strategies in the population began to evolve according to their payoffs. As a result, the strategy of unconditional defection was always evolutionary stable (Hofbauer & Sigmund, 1998) while pure cooperators could not survive. Nevertheless, the Tit-for-Tat strategy is evolutionary stable as well, which promotes cooperation based on reciprocity (Axelrod & Hamilton, 1981).

Recently, evolutionary dynamics in structured populations has attracted much attention, where the structured population denotes an infinite and well-mixed population which simplifies the analytical description of the evolution process. In real populations, individuals are more likely affected by their neighbors than those who are far away, but the spatial structure of population is omitted in the iterated PD game. To study the spatial effects upon strategy frequencies in the population, (Nowak & May, 1993) have introduced the spatial PD game, in which players are located on the vertices of a two-dimensional lattice, whose edges represent connections among the corresponding players. Instead of playing with all other contestants, each player only interacts with his neighbors. Without any strategic complexity the stable coexistence of cooperators and defectors can be achieved. However, the model presented in (Nowak & May, 1993) assumes a noise free environment. To characterize the effect of noise, (Szabó & Toke, 1998) have presented a stochastic update rule that permits irrational choice. Besides, (Perc & Szolnoki, 2008) account for social diversity by stochastic variables that determine the mapping of game payoffs to individual fitness. Furthermore, many other works centered on the lattice structure have also been done. For example, (Vukov & Szabó, 2005) have presented a hierarchical lattice and shown that for different hierarchical levels the highest frequency of cooperators may occur at the top or middle layer. For more details about evolutionary games on graphs, see (Doebeli & Hauert, 2005; Nowak, 2006; Szabó & Fath, 2007) and references therein.

Yet, as imitations of real social networks, the evolutionary game on lattices assumes that there is a fixed neighborhood for each player. Nevertheless, this assumption does not always hold for most of real social networks. Unlike models mentioned above, the relationships among players (data points) in our model are represented by a weighted and directed network, which means that players are not located on a regular lattice any more. And the network will evolve over time because each player is allowed to apply an edges-removing-and-rewiring (ERR) function to change his connections between him and his neighbors. Furthermore, the payoff matrix of any two players in the proposed model is also time-varying instead of a constant payoff matrix, for instance, the payoff matrix in PD game. As a consequence, when the evolutionarily stable strategies emerge in the network, it will be observed that only a few players (data points) receive considerable connections, while most of them have only one connection. Naturally, players (data points) are divided into several parts (clusters) according to their evolutionarily stable strategies.

3. Proposed model

Assume a set $\boldsymbol{X}$ with N players, $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N\}$, which are distributed in a m-dimensional metric space. In this metric space, there is a distance function $d : \boldsymbol{X} \times \boldsymbol{X} \to \mathbb{R}$, which satisfies the condition that the closer any two players are, the smaller the output is. Based on the distance function a distance matrix is computed whose entries are distances between any two players. Next, a weighted and directed k-nearest neighbor (knn) network, $G_0(\boldsymbol{X}, E_0, d)$, is formed by adding k edges directed toward its k nearest neighbors for each player, which represents the initial relationships among all players.

Definition 1. If there is a set $\boldsymbol{X}$ with N players, $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N\}$, the initial weighted and directed knn network, $G_0(\boldsymbol{X}, E_0, d)$, is created as below.