Pattern Recognition Letters 116 (2018) 114–120

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A solution to the classification problem with cellular automata

Arif Orhun Uzun[a], Tuğba Usta[a], Enes Burak Dündar[b], Emin Erkan Korkmaz[a,*]

[a] Yeditepe University, İstanbul, Turkey
[b] Boğaziçi University, İstanbul, Turkey

ARTICLE INFO

Article history:
Received 14 April 2018
Available online 9 October 2018

Keywords:
Classification
Cellular automata
Heat Transfer
Big Data

ABSTRACT

Classification is the task of labeling data instances by using a trained system. The data instances consist of various attributes and in order to train the system, a set of already labeled data is utilized. After the training process, success rate of the system is determined with separate test sets. Various machine learning algorithms are proposed for the solution of the problem. On the other side, Cellular Automata (CA) provide a computational model consisting of cells interacting with each other based on some predetermined rules. In this study, a new approach is proposed for the classification problem based on CA. The method maps the data instances in the training data set to cells of an automaton based on the attribute values. When a CA cell receives a data instance, this cell and its neighbors are heated based on a heat transfer function. A separate automaton is heated for each class in the data set and hence a characteristic heat map is obtained for each class at the end of the procedure. Then new instances are classified by using these heat maps. The success rate of the algorithm is compared with the results of other known classification algorithms in the experiments carried out.

1. Introduction

Classification is a supervised learning process in which a set of labelled data is utilized to partition the instance space. If the classification model performs well on the training data utilized, then it becomes possible to predict appropriate labels for unseen data instances. For classification tasks, data instances are represented in the following format, $\{(x_1, l_1),(x_2, l_2),\ldots ,(x_k, l_k)\}$, where x_i is a data instance with n attributes and l_i is the class label of the data instance. The classification task can be defined as deducing the function f: $X \rightarrow L$ where X is the instance space and L is the output space. Accuracy of the function f is calculated with a separate test set. Many different techniques are used for acquiring the function f such as Naive Bayes, support vector machines and decision trees [7,13,18].

Cellular Automata (CA) provide a means for computation based on the interaction of interconnected cells in a discrete system. The interaction is carried by using some predetermined rules. Most popular CA application is Conway's Game of Life [8]. Other than that, CA is mostly used as a simulation tool for various disciplines [2,23].

In this study, a model based on CA is proposed as a solution to the classification problem. For any data set with n attributes and c classes, c two dimensional (2D) automata are utilized in the computation where a 2D-automaton is reserved for each class. Each attribute a_i is associated with a row of this 2D-automaton. Hence, each automaton would have n rows and m columns, where m is parameter that can be set to different values in the algorithm. Then the automaton cells are heated based on the attribute values of the instances belonging to the class associated with this automaton. The warming procedure is repeated for each automaton in the model by using the instances of the classes corresponding to these automata. A characteristic heat map is obtained for each class in the dataset at the end of the procedure. Then the new instances can be classified by using these heat maps belonging to different classes.

Performing the classification task by using these characteristic heat maps is the novel approach proposed in this study. The method is quite simple and is advantageous in terms of runtime performance compared to other classification methods. The details of the warming and the classification procedures utilized in the study can be found in Section 3.

2. Related work

In the literature, various approaches exist where CA is utilized for the classification task. For instance, in [19], a novel method named as Classificational Cellular Automata(CCA) has been pro-

* Corresponding author.
E-mail addresses: burak.dundar@boun.edu.tr (E.B. Dündar), ekorkmaz@cse.yeditepe.edu.tr (E.E. Korkmaz).

https://doi.org/10.1016/j.patrec.2018.10.007
0167-8655/

(a) Moore (b) Von Neumann

Fig. 1. Neighborhood Types.

posed. CCA combines various classification techniques as a Multi Classifier System. In [5], data classification task has been solved where the relationship between the data attributes are extracted by utilizing a cellular learning automaton. In [15], the classification process is again carried out by a CA. An ensemble of classifiers is built by a CA in the study. Each cell of the CA keeps information about the success of the corresponding classifier. Then, a set of transaction rules determine the energy in each cell by using the information kept in it. Learning process is conducted by the interaction of neighbourhood cells based on these transaction rules. For each dataset, the performance may depend on parameters like the size of CA and transaction rule parameters. Fawcett [6] is another example where the classification problem is solved by using CA. The study focuses on the question of how data points with real-valued vectors should be mapped into a discrete space. Moreover, the method proposed uses a voting rule for determining the class of a selected cell.

Also, in [20], a CA model utilizing a Gaussian kernel has been proposed in order to solve the classification problem. Yet, the method is analysed only on 2-dimensional data, not tested on well-known datasets, and not compared with well-known classification methods. Cellular Automata have been widely utilized for simulation, too. In [1], a CA approach has been proposed for the problem of heat conduction. In [14], the simulation of the temperature changes on the earth surface is carried out by cellular automata. On the other side, in [3], simulations of heat and mechanical energy have been demonstrated on cellular automata. These approaches indicate that various real world processes can be simulated via cellular automata including heat conduction.

On the other hand, in [4], clustering problem has been solved by utilizing a heat propagation procedure in an n-dimensional CA framework where n is the number of attributes in the data. The main advantage achieved against the well-known clustering algorithms is that no distance metric is needed in the proposed algorithm. Therefore, clustering process becomes easier for big datasets. However, space complexity of the method proposed in [4] is $O(m^n)$ where m is the number of cells in each dimension and n is the number of attributes. Therefore the method is applicable to low dimensional datasets. In this study, a similar procedure with [4] is utilized, but operation is carried out on two dimensional CA. Hence, the method is applicable to high dimensional datasets, too.

3. Methodology

Cellular Automata provide a framework in which each cell interacts with neighbour cells based on a set of predefined rules. There are two common neighbourhood relationships utilized: the Moore and Von Neumann which are shown in Fig. 1. In n-dimensional space, Von Neumann neighbourhood has $2n$ neighbours, the Moore has $3^n - 1$. In this study, linear neighborhood is utilized in 2D-CA where each cell interacts with only two neighbor cells in the same row.

As noted in Section 1, a separate 2D-CA is utilized for each class in this study. Thereby, characteristics of each class is obtained af-

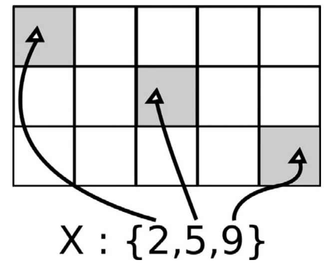

Fig. 2. Mapping procedure.

Algorithm 1: Overall algorithm of the training process.

```
Function Train(D_train,n,c,CA):
    D_train: The training dataset
    n  : Number of attributes in the dataset
    c  : Number of classes in the dataset
    CA : An array of c Cellular Automata
    foreach data_instance ∈ D_train do
        label=data_instance[n + 1]
        Map-Data(data_instance, n, CA[label]);
    end
    for i ← 1 to c do
        Set-Heat(CA[i]);
        Distribute-Heat(CA[i]);
    end
Function Map-Data (d,n,A):
    for i ← 1 to n do
        index = FindIndex(i, D_i);
        // Increment the counter of the corresponding
           cell in A
        A[i][index]_counter + +;
    end
Function Set-Heat(A):
    foreach cell ∈ A do
        cell_temp = ln(cell_Counter + 1)
    end
Function Distribute-Heat(A):
    Generate-Heat-Waves(A);
    while A has heat to distribute do
        foreach cell ∈ A do
            if cell has incoming_heat then
                IncreaseHeat(cell, incoming_heat);
                TransferHeat(cell, incoming_heat);
            end
        end
    end
```

Table 1
Datasets utilized in the experiments.

Data set	# of Data instances	Attributes	Classes
Australian	690	14	2
Banknote	1372	4	2
Breast-Wisconsin	699	9	2
Glass	214	9	6
Haberman	306	3	2
Heart	270	13	2
Iris	150	4	3
Pima	768	8	2