

Autômatos celulares para reconhecimento de padrões

Hallefy Ferreira¹, Núbia Rosa da Silva¹

¹Instituto de Biotecnologia – Universidade Federal de Goiás (UFG)
Av. Dr. Lamartine Pinto de Avelar, 1200 – 75704-020 – Catalão – GO – Brazil

{hallefy.fc,nubiasrosa}@gmail.com

Abstract. *Cellular automata has been used in several tasks, such as the classification task. This is because the ease of implementation, and based on few rules we can obtain great results, as compared with other classification methods. This paper has the objective utilize cellular automata to classify the dataset IRIS, and compare their results with an algorithm widely used in the classification task, KNN. Our proposed method obtained a classification rate of 96,6%.*

Resumo. *Autômatos celulares vem sendo utilizados em várias tarefas, como por exemplo, classificação. Isso acontece devido a facilidade de implementação, e a partir de poucas regras pode-se obter bons resultados, quando comparado a outros métodos. Este trabalho tem como objetivo utilizar autômatos celulares para efetuar a classificação da base de dados IRIS, e comparar seus resultados com um algoritmo muito usado na tarefa de classificação, o KNN. O algoritmo proposto no artigo obteve uma taxa de classificação de 96,6%.*

1. Introdução

Um autômato celular (CA – *Cellular Automata*) é um modelo matemático discreto com uma população de células que evolui a cada geração a partir de uma regra de transição [Chopard and Droz 1998]. John Horton Conway, um matemático britânico, desenvolveu em 1970 um importante trabalho para a época, quando se fala em autômato celular o exemplo mais comum é o jogo da vida, desenvolvido por ele [Adamatzky 2010]. No jogo da vida uma célula morre se possui menos de dois vizinhos ou se possui mais de três vizinhos. Uma célula morta com três vizinhos se torna uma célula viva e se uma célula viva possui exatamente dois ou três vizinhos, a célula continua viva.

Autômatos celulares estão presentes em diversas áreas como biologia, finanças, processamento de imagens, computação paralela, simulações de situações reais, como por exemplo, determinar a forma como um certo tipo de vírus pode se espalhar em determinada região, simular a propagação de um incêndio em florestas ou realizar um estudo sobre a expansão dos centros urbanos [Chandra and Vidushi 2012, Fu and Milne 2003, Tian et al. 2016]. Neste contexto, o principal objetivo deste trabalho é empregar os conceitos de autômatos celulares para reconhecimento de padrões, classificando amostras de acordo com suas características. Pode-se definir classificação como a tarefa de classificar ou dispor dados em determinadas classes, através da extração de informação ou conhecimento de uma base de dados. A metodologia consiste em analisar os vizinhos da célula para realizar a tarefa de classificação. Uma modificação muito importante foi feita na regra de Moore para que o autômato considere o seu estado atual para a classificação pois o estado atual da célula possui alta relevância para definir o seu próximo estado. Para verificar a potencialidade da metodologia, a base de dados IRIS [Lichman 2013] será utilizada nos experimentos.

2. Autômatos celulares

De acordo com John von Neumann [Von Neumann et al. 1966], um autômato celular pode ser definido como um modelo capaz de se auto-reproduzir a partir de regras definidas.

Definição I. Um autômato é formado pela 5-tupla $(\alpha, \beta, \gamma, \delta, \Theta)$, onde:

- (i) α é um conjunto finito de estados;
- (ii) β é um conjunto finito de símbolos;
- (iii) γ é uma função de transição;
- (iv) δ é o estado inicial do autômato, antes de ser aplicada uma função de transição sobre determinada célula, onde $\delta \in \alpha$;
- (v) Θ é um conjunto de estados de α , chamados também de estados de aceitação.

Em CA existem algumas definições de vizinhos, as mais utilizadas são as definições de vizinhança de von Neumann e Moore [Wolfram 1983]. Em uma tesselação 2d, a definição de von Neumann consiste em verificar os estados de suas células vizinhas que estão ao norte, sul, leste e oeste, já a definição de Moore acrescenta a possibilidade de verificar o estado das células que estão a nordeste, sudeste, sudoeste e noroeste. Para o desenvolvimento do CA utilizado neste artigo, a definição de Moore foi indicada por obter uma boa precisão na tarefa de classificação.

3. Autômatos celulares como classificador

O algoritmo desenvolvido tem como base as análises feitas em [Adwan et al. 2013, Fawcett 2008], onde verifica-se o estado de cada célula na vizinhança de Moore para classificar a célula atual como pertencente a uma determinada classe. Neste trabalho, todos os vizinhos incluindo o centro são considerados para a mudança de estado da célula para uma transição do tempo t para o tempo $t + 1$. Logo, para classificar o estado das células na próxima geração, uma regra de transição é definida para determinar como o autômato irá se comportar.

- Regra de transição de cada célula:
 - 0 : Se a soma dos vizinhos de cada classe for igual;
 - A : Se o número de vizinhos de classe A é maior;
 - B : Se o número de vizinhos de classe B é maior;
 - C : Se o número de vizinhos de classe C é maior;
 - X : Se o número de vizinhos são iguais.

4. Experimentos e resultados

A base dados utilizada para a realização do experimento foi a base de dados IRIS [Lichman 2013], que possui 3 classes de flores, sendo elas: setosa, versicolor e virgínica (A , B e C , respectivamente). Cada classe possui 50 instâncias resultando em 150 instâncias. Cada flor é composta por 4 características: comprimento da pétala, largura da pétala, comprimento da sépala e largura da sépala. O CA foi implementado na linguagem C em duas dimensões. Devido a base de dados IRIS possuir quatro características, duas matrizes bidimensionais foram construídas, onde uma matriz possui relação com as características da pétala e a outra matriz com características da sépala. Para realizar a classificação foram utilizadas as duas matrizes, onde se verifica o número de vizinhos de cada célula e a partir da regra de transição é feita a classificação. As Figuras 1(a) e 1(b) representam a distribuição das classes que foram classificadas pelo autômato desenvolvido.

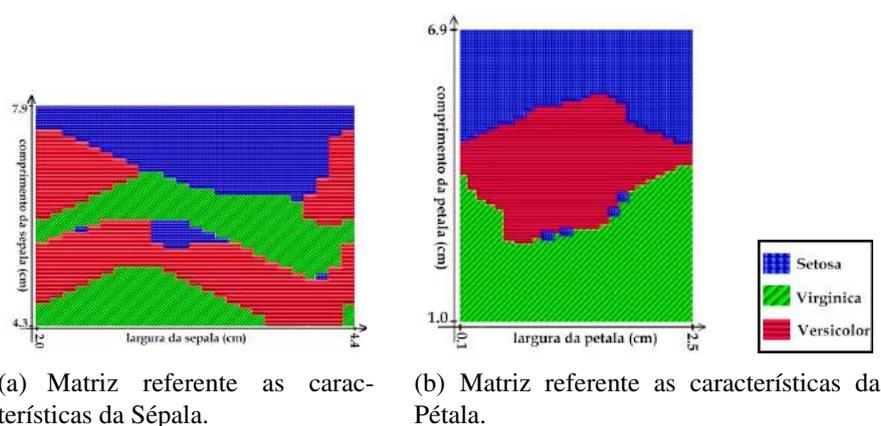


Figure 1. Distribuição das classes de acordo com o CA desenvolvido utilizando em (a) as características da sépala e em (b) as características da pétala.

O conjunto de dados foi dividido em treinamento e teste utilizando a técnica *k-fold cross-validation*, que consiste em dividir o conjunto em *k folds*, onde cada *fold* contém um número balanceado de amostras de cada classe. A base de dados IRIS possui 50 amostras para cada uma de suas três classes de espécies de flores, resultando em 10 folds contendo 15 amostras cada. Com o conjunto de treinamento preparado, um *fold* é selecionado para ser o conjunto de teste e os outros $k - 1$ *folds* formarão o conjunto de treinamento, no qual o autômato é treinado. Esse procedimento é feito para cada um dos folds. O método KNN foi utilizado para comparação dos resultados juntamente com o algoritmo proposto em [Ahangaran et al. 2017] denominado de *Cellular learning automata* (CLA), pois utilizam conceitos semelhantes de vizinhança para determinar um novo estado para a célula, o algoritmo CLA obteve 95% de precisão. Os parâmetros para o KNN foram $K = 1$ e distância euclidiana.

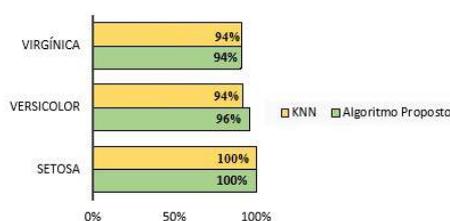


Figure 2. Resultados ao comparar a metodologia proposta e o método KNN.

Ao utilizar a metodologia proposta, uma taxa de classificação de 96,6% foi atingida (Figura 2). O algoritmo de classificação *k*-Nearest Neighbor (KNN) obteve 96% de acerto utilizando o mesmo esquema de validação. Nota-se que o modelo proposto conseguiu melhores resultados do que o KNN onde na classe setosa a taxa de classificação ficou igual em ambos algoritmos: 100%. Já nas outras classes versicolor e virgínica, houve uma diferença na taxa de classificação, onde o algoritmo proposto neste artigo se mostrou superior, obtendo uma taxa de classificação na classe versicolor de 96% contra 94% do KNN, e na classe virgínica obteve uma taxa de classificação de 94%, superando os 94% do KNN.

5. Conclusões

O objetivo deste trabalho foi desenvolver um autômato celular que pudesse realizar tarefas de classificação, para isso, a base de dados IRIS foi utilizada para realizar os experimentos e com os resultados, pode-se perceber que autômatos celulares possuem uma vasta gama de aplicação seja na área de ciência da computação como biologia, economia e áreas afins. Como projetos futuros temos como objetivo fazer a fusão de mineração de dados e autômatos para conseguir extrair informações, comportamentos ou regras de um determinado conjunto de dados. De um modo geral pode-se dizer que a tarefa de classificação é um tipo de mineração de dados no qual a partir de uma regra, classifica-se um conjunto de dados. A principal vantagem de se utilizar o CA proposto está na facilidade de implementação, além de ser uma abordagem nova e bastante promissora. É importante destacar que podem ser utilizados dados com mais de 2 dimensões (números de características), porém estes dados devem ser quebrados em matrizes bidimensionais para não contrariar a definição de vizinhança utilizada neste trabalho.

References

- Adamatzky, A. (2010). *Game of life cellular automata*, volume 1. Springer.
- Adwan, O., Huneiti, A., Ayyal Awwad, A., Al Damari, I., Ortega, A., Abu Dalhoum, A. L., and Alfonseca, M. (2013). Utilizing an enhanced cellular automata model for data mining. *International Review on Computers and Software*.
- Ahangaran, M., Taghizadeh, N., and Beigy, H. (2017). Associative cellular learning automata and its applications. *Applied Soft Computing*, 53:1–18.
- Chandra, M. A. and Vidushi (2012). Article: A study and analysis on cellular automata based classifier in data mining. *IJCA Proceedings on International Conference on Advances in Computer Applications 2012*, ICACA(1):30–35. Full text available.
- Chopard, B. and Droz, M. (1998). *Cellular automata*. Springer.
- Fawcett, T. (2008). Data mining with cellular automata. *ACM SIGKDD Explorations Newsletter*, 10(1):32–39.
- Fu, S. and Milne, G. (2003). Epidemic modelling using cellular automata. In *Proc. of the Australian Conference on Artificial Life*. Citeseer.
- Lichman, M. (2013). UCI machine learning repository.
- Tian, G., Ma, B., Xu, X., Liu, X., Xu, L., Liu, X., Xiao, L., and Kong, L. (2016). Simulation of urban expansion and encroachment using cellular automata and multi-agent system model – a case study of tianjin metropolitan region, china. *Ecological Indicators*, 70:439–450.
- Von Neumann, J., Burks, A. W., et al. (1966). Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1):3–14.
- Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of modern physics*, 55(3):601.