Cognitive Consistency Models Applied to Data Clustering

Thales Vaz Maciel[1,2](✉) and Leonardo Ramos Emmendorfer[2]

[1] Sul-rio-grandense Federal Institute for Education, Science and Technology, Bagé 96418-400, Brazil
thalesmaciel@ifsul.edu.br

[2] Center for Computational Science, Federal University of Rio Grande, Rio Grande 96203-900, Brazil
leonardoemmendorfer@furg.br

Abstract. Data clustering concerns the discovery of partitions in data such that items from the same groups are as similar to each other as possible, and items from different groups are as dissimilar to each other as possible. The literature presents vast diversity on data clustering approaches, including systems that model the behavior of social individuals from different species. This work proposes a clustering algorithm that is reasoned upon the social theory of cognitive dissonance and the psychological theory of balance. We investigate whether psychological balance aware decision-making capabilities would affect the data clustering task. Partial results revealed the superiority of the proposed approach according to 5 out of 9 clustering quality metrics, when compared to other clustering algorithms.

Keywords: Data clustering · Cognitive dissonance · Balance theory

1 Introduction

Data clustering is defined as the discovery of partitions in datasets, in the context of Machine Learning. The intuitive idea behind the notion of data clustering is the search for groups of objects which share some kind of similarity [7]. The resulting partitions (or clusters) contain elements that are as similar as possible to those that are in the same cluster and also as dissimilar as possible to those elements in other clusters. The task of clustering is also relevant due to the ability to reveal useful patterns from datasets. The topic has gained importance over recent years, mostly as a result of successful applications in a wide range of fields, such as biology, medicine, psychology, image processing, among others [11].

There is a large number of clustering algorithms available in the literature, which tend to specialize in specific characterizations of data [3,8]. The diversity in clustering strategies is very prominent. Centroid-based strategies, for instance, recurrently self-adapt until they meet stability criteria to the detection of data density decays that potentially translate into cluster borders.

L. Rutkowski et al. (Eds.): ICAISC 2021, LNAI 12855, pp. 183–191, 2021.
https://doi.org/10.1007/978-3-030-87897-9_17

Biologically-inspired models should also be mentioned, where the natural clustering behavior of individuals from certain species can be simulated computationally.

Clustering algorithms that are inspired by the social behavior of individuals from different species are described in the literature, including, for instance, the impact of conflicting cognitions on the individual's behavior. One can notice, however, a lack of the modeling of certain aspects of psychological processes that guide the behavior of individuals who compose society.

The theory of cognitive dissonance, proposed by Leon Festinger in 1957 [4], refers to a psychological phenomenon caused by situations where conflicting attitudes, beliefs or behaviors are present in the individual's psyche. According to the theory, such sort of situation causes mental discomfort, which leads the individual to alter one of the conflicting cognitions in order to reduce the discomfort and thus restore its psychological balance. Heider (1957) [6] presented the balance theory, which explains changes of attitude in individuals that seek psychological balance. According to the author, individuals exist in a constant need for maintenance of their own cognitive consistency, solving or avoiding conflicting values, beliefs, and situations that relate to the environment they are placed. Relations between two individuals and from those towards a general object, which may also be an equally cognitive-capable third individual, are represented in triads [6]. In a triad, the product of the three positive or negative relations represents the psychological state of the situation as a whole, either balanced or unbalanced.

Cartwright and Harary (1956) [1] proposed an extension of Heider's model that is based on graph theory, thus enabling the simultaneous analysis of multiple relations, whereas the original model considered strictly 3 entities. One of the authors' contributions to balance theory was the exact possibility of representing the relations between an unlimited set of entities. In that sense, Flament (1963) [5] presented a theorem stating that a signaled complete graph is balanced only if all possible triads, from the relations that compose the graph, are balanced. However, in a more profound application of graph theory to interpersonal relationships situation, the author revised their previous statement to be insufficient due to a consideration that different individuals may have different perceptions over the situations they are involved.

This work proposes a novel clustering algorithm which relies on theories of cognitive dissonance [4] and psychological balance [1,6] as the fundamental mechanisms responsible for group formation. The solution herein proposed works as follows: each data object is assigned to a single agent, which interacts with other agents similarly generated in order to iteratively achieve a stable group formation, where neighboring agents agree about group formation. Each agent senses its environment and reacts, at each iteration, aiming to reduce its own cognitive discomfort resulting from eventual disagreements with its neighbors.

The aim of this work is to investigate whether the adoption of human cognitive models related to cognitive consistency can be useful for data clustering, leading to increased quality of the discovered clusters. Specifically, we aim to define an original clustering algorithm which regulates the decision making processes related to the clustering task by using concepts from cognitive dissonance and balance theories.

This work is organized as follows: Sect. 2 introduces recent clustering algorithms that apply conflicting cognitions at some level. Section 3 presents the proposed method. In Sect. 4 the results from experiments are shown. Section 5 concludes the paper.

2 Related Work

This section presents a review of clustering algorithms that relate to cognitive consistency inherent concepts and that apply reasoning based on such concepts to the data clustering logic.

Cohen and Castro (2006) [2] presented a clustering algorithm based on the particle swarm optimization (PSO) algorithm, called Particle Swarm Clustering (PSC). As a clustering strategy, the authors modeled the human tendency of adapting their behavior according to the influence of the environment by minimizing the differences in opinions and ideas through time, but also considering the individual's past experiences as emergent behavior.

These ideas are closely related to the concepts of cognitive dissonance [4] and balance theory [6] as they propose that conflicting ideas may be reorganized by actions of an individual and also that these actions as performed in a search for psychological balance.

In recent years, there have been algorithms that share those basic premisses and develop diversely from then on. Examples are the Ant Colony Optimization Clustering with Chaos (ACOCC) [12], Swarm Clustering Algorithm (SCA) [13] and Fully-Informed Particle Swarm Clustering (FIPS) [9].

Yang et al. (2018) [12] proposed an improvement to the ACO Clustering (ACOC) by including a chaotic function in the initialization and feromone update phases so that it produces disturbances in the ants' movements in the systems so that it leads them to find a globally optimized solution instead of a local one.

Zhu et al. (2018) [13] proposed an improvement to general PSC-based clustering algorithms, suggesting that existing related algorithms had been designed so that the swarm particles represented cluster centers. The contribution was that, instead, swarm particles would represent each data instance individually and that these particles would recurrently interact and movement towards more densely populated areas in the systems, therefore aggregating to for clusters.

Mansour and Ahmadi (2019) [9] proposed a PSO-based clustering algorithm with the hybrid application of different neighborhood topologies, namely total, ring, and distance-based. Authors considered that their contribution was in the use of the latest topology as it considers a certain number of closest particles in the clustering process and that neighborhoods change in each iteration.

Godois et al. (2018, 2020) [7] proposed a clustering algorithm that considers every data instance as an agent of a multiagent system. In the system, the agents have a multidimensional view circle, which is divided into sectors and, in each iteration, each agent would choose to move towards the center of the view circle sector that is most dense, in terms of population by other agents. This occurs recurrently in a way that the agents can change their assigned cluster in each iteration until no cluster changes are performed in a complete iteration. The difference between the first and the latest is that the first proposed that, after choosing a sector of their view circle towards which an agent would move in an iteration, it would assume the same cluster as the other agent from the chosen sector that is closest to it, while, in the latest, the assumed cluster would be the most representative from the chosen sector.

3 Cognitive Dissonance Clustering Algorithm

This section presents a novel algorithm, which is called Cognitive Dissonance Clustering (CDC). We have shown how cognitive dissonance can drive the clustering process.

The algorithm is based on the movement of agents which locally search for a better cognitive environment, where each agent represents a data point from the input dataset. An agent senses its environment, within a limited vision scope around its current coordinates. This resulting circumference is then divided into equally distributed sectors. Figure 1 illustrates the vision scope of an agent, which is split into 8 sectors.

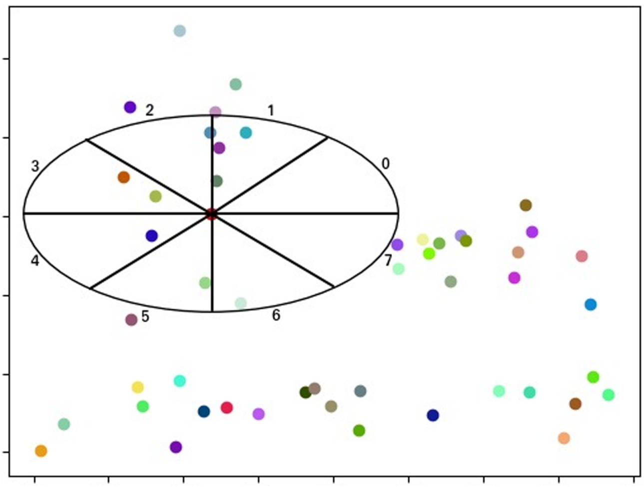

Fig. 1. The vision scope of an agent, which is split into 8 sectors labeled as 0,1, $\cdots$, 7.

Agents' decision making is parametrized. The parameters are, namely, the size of the step each agent takes in each iteration, the size of the radius of the agents' perception circle which limits the space in which each agent can perceive other agents, the number of sectors to which the agents' perception circles are divided (these are referred to as perception sectors) and the cognitive satisfaction threshold for the agents, which defines the psychological balance degree from which an interagent situation is classified as positive or negative. Figure 2 illustrates the algorithm proposed.

Algorithm 1 Cognitive Dissonance Clustering (CDC)

Input:
A dataset $\Delta = \{ \delta_1, \delta_2, \cdots, \delta_n \}$.
Vision radius Γ.
Step size λ.
Number of sectors ϕ around each agent.
Output:
A set of partitions P.

for all $\delta_{\mathrm{i}} \in \Delta$ **do**
    Create an agent $\alpha_i \in A$.
    Set the coordinates of α_i as the same of data point δ_i.
    Set the cluster label of α_i as i.
end for
while the partitioning did not converge **do**
    Ψ = set of positive and negative relations between agents in A
    for all $\alpha_i \in A$ **do**
        Divide the perception circle of α_i into ϕ sectors, $\Omega = \{ \omega_0, \omega_1, \cdots, \omega_{\phi-1} \}$.
        $\omega^\star \leftarrow$ select the best $\omega_i \in \Omega$, which maximizes psychological comfort of ${\alpha_i}^\dagger$.
        Move α_i by λ towards the centroid of $\omega^\star$.
        Set the cluster label of α_i to the the prevalent cluster label in $\omega^{\star\dagger}$.
    end for
end while
for all $\delta_{\mathrm{i}} \in \Delta$ **do**
    $\delta_{\mathrm{i},\kappa} \leftarrow \alpha_{\mathrm{i},\kappa}$.
end for
†Ties are broken at random.

Fig. 2. Cognitive Dissonance Clustering (CDC).

At the beginning of each iteration, an interagent relation graph is built which links each agent to every other in the system. Initially, a positive relationship is established between each agent and its closest counterpart and also to all other agents that are at that same smallest distance to it. Figure 3 illustrates shows all positive relations at the first iteration for an illustrative dataset. The relation to the other agents is considered as negative, initially. At each iteration, the agent moves to the sector where it feels more comfortable, according to the psychological models proposed in [6] and [1]. The iterative process is described as follows.

Conflicting relationships between pairs of agents might appear. For instance, agent A might have a positive relationship with B but B has a negative relationship with A. In those cases, the positive relationship prevails. Reciprocal relations, for example, when an agent is the closest to another and vice-versa or

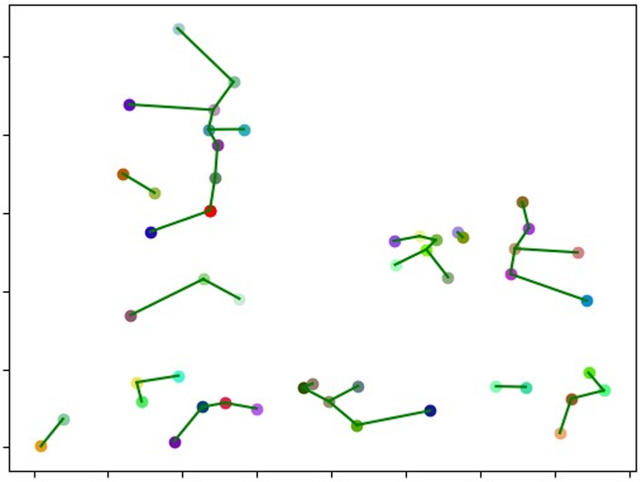

Fig. 3. Positive relations at the first iteration are represented as green links. (Color figure online)

an agent is not the closest to another and vice-versa, are just respectively defined as positive or negative. Each agent then verifies the psychological balance degree in each of its perception sectors at the moment.

In the process of verifying the balance degree of an agent's perception sector, only the interagent relations that are completely inside the sector are considered. From those, according to [1], relation triads are calculated as the product of the relations that compose the found relation triads. In order to discover the psychological state of the triads in each sector in which densities are at least 3, positive relations are labeled as +1 and negative relations are labeled as −1 [6]. Moreover, the product of the relations that form a triad corresponds to the psychological balance of the triad as a whole. This enables a triad to be evaluated as either positive or negative.

After defining the balance of every possible relation triad in an agent's perception sector, it is viable to calculate the balance of the sector as a whole. According to [5], a complete signaled graph, such as presented for verifications in the proposed algorithm, may be classified according to its balance degree. Balance degree calculation is applied in this context to define the psychological state for each sector of each circle of perception of each agent in the system. Later on, those sector states are summarized as either positive or negative, respectively, if the sector's balance degree reaches a defined cognitive satisfaction threshold, which is shared amongst every agent in the system.

Each agent then decides about the direction towards which it is to move at the current iteration. Agents decide to move towards the direction of the sector that leads to higher psychological comfort for itself. In the case of an agent detecting the same level of psychological comfort in more than one direction or when all the directions present the same neutral balance, the agent chooses to move towards the sector that presents the highest populational density at the moment. After performing its movement towards a sector, the agent assumes the cluster that represents the majority of the agents in that sector.

The algorithm runs iteratively until no agents change their cluster labels, or a threshold in the number of iterations is achieved. The final set of cluster labels represents the resulting partitioning from the algorithm.

4 Results and Discussion

This section describes the obtained results for the proposed clustering method, in terms of the quality of the resulting models for a specific dataset. Performance evaluation is assessed in comparison to other 3 data clustering algorithms, namely K-means, DL2, and AVC [7]. Except for K-means, which is applied here strictly as a comparative baseline, DL2 and AVC relate to the proposed algorithm in a more considerable level.

The performance metrics that were used to evaluate the clusters' quality in the discovered models where the adjusted Rand index (ARI), normalized mutual information (NMI), homogeneity, completeness, V-measure, and the coefficients of Fowles-Mallows (FM), Silhouette, Calinski-Harabasz (CH) and Davies-Bouldin (DB) according to definitions provided in [10].

The bidimensional dataset that was used in the tests was assembled specifically for the purpose of this comparison. It is composed of 50 instances, arranged over 3 classes that represent natural clusters with varying shapes, sizes, and densities. In total, each algorithm was executed 50 times using the dataset and the average values for each performance metric obtained from the discovered models are shown in Table 1.

Table 1. Average quality indicators for the partitionings discovered by K-means, DL2, AVC and CDC.

Average Performance Metric	K-means	DL2i	AVC	CDC
Adjusted Rand Index	0.481	0.788	0.845	<u>0.932</u>
Normalized Mutual Info	0.570	0.839	0.873	<u>0.924</u>
Homogeneity	0.546	0.930	0.928	<u>0.943</u>
Completeness	0.660	0.780	0.840	<u>0.915</u>
V-Measure	0.586	0.847	0.879	<u>0.928</u>
Fowles-Mallows	0.635	0.856	0.895	<u>0.955</u>
Silhouette	0.486	<u>0.532</u>	0.514	0.429
Calinski-Harabasz	55.966	<u>68.794</u>	62.557	45.154
Davies-Bouldin	0.737	<u>0.565</u>	0.606	0.663

As shown in Table 1, the CDC algorithm was superior on average to K-means, DL2, and AVC in 6 of the 9 cluster quality metrics that were applied in the cluster analysis. Namely, ARI, NMI, homogeneity, completeness, V-measure, and

FM were the clustering evaluation methods that better evaluated the models discovered by CDC and proposed that it is first-rated over the other 3 benchmarked algorithms. On the other hand, the coefficients of Silhouette, CH, and DB did not corroborate with the CDC superiority hypothesis, as it was only able to outperform K-means, according to those two measures.

From the results in Table 1, it is also possible to infer that CDC was superior to K-means, DL2, and AVC according to all internal evaluation metrics used. Even though competitive, CDC wasn't first-rated on average for the performance in the described tests according to the external cluster evaluation metrics adopted.

5 Conclusion

This paper revealed that cognitive consistency models, which are useful for understanding human behavior, can be adopted for the data clustering task. A novel algorithm was proposed, which mimics the response of actual human individuals under cognitive discomfort.

Initial results are promising, as shown by quantitative evaluation of the partitioning obtained. Further investigation should consider a wider range of study cases and datasets. Other types of cognitive models could also be adopted in the future.

References

1. Cartwright, D., Harary, F.: Structural balance: a generalization of Heider's theory. Psychol. Rev. **63**(5), 277 (1956)
2. Cohen, S.C., de Castro, L.N.: Data clustering with particle swarms. In: 2006 IEEE International Conference on Evolutionary Computation, pp. 1792–1798. IEEE (2006)
3. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: Density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD-96 Proceedings, pp. 226–231. AAAI Press (1996)
4. Festinger, L.: A Theory of Cognitive Dissonance. vol. 2, Stanford University Press, Redwood (1957)
5. Flament, C.: Applications of Graph Theory to Group Structure. Prentice-Hall, Hoboken (1963)
6. Fritz, H., et al.: The psychology of interpersonal relations (1958)
7. Godois, L.M., Adamatti, D.F., Emmendorfer, L.R.: A multi-agent-based algorithm for data clustering. Prog. Artif. Intell. **9**(4), 305–313 (2020). https://doi.org/10.1007/s13748-020-00213-3
8. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297. University of California Press, Okland(1967)
9. Mansour, E.M., Ahmadi, A.: A novel clustering algorithm based on fully-informed particle swarm. In: 2019 IEEE Congress on Evolutionary Computation (CEC), pp. 713–720 (2019)

10. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
11. Xu, R., Wunsch, D.C.: Clustering. Wiley-IEEE Press, Piscataway (2009)
12. Yang, L., Li, K., Zhang, W., Ke, Z., Xiao, K., Du, Z.: An improved chaotic ACO clustering algorithm. In: 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 1642–1649 (2018)
13. Zhu, W., Luo, W., Ni, L., Lu, N.: Swarm clustering algorithm: let the particles fly for a while. In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1242–1249 (2018)