The Thirty-First International Florida
Artificial Intelligence Research Society Conference (FLAIRS-31)

A Case-Based Reasoning and Clustering Framework for the Development of Intelligent Agents in Simulation Systems

Marcos R. B. Lucca,[1] Alcides G. Lopes Junior,[1,2] Luis A. L. Silva,[1] Edison P. Freitas[2]

[1]Graduate Program in Computer Science, Federal University of Santa Maria, Av. Roraima, 1000, Santa Maria, 97105-900, Brazil.
marcosrlucca@gmail.com, ajunior@inf.ufsm.br, luisalvaro@inf.ufsm.br

[2]Graduate Program in Computer Science, Federal University of Rio Grande do Sul, CP 15064, Porto Alegre, 91501-970, Brazil.
epfreitas@inf.ufrgs.br

Abstract

Artificial Intelligence (AI) techniques are essential to the modeling of realistic behaviors for agents in simulation systems. Although Case-Based Reasoning (CBR) and Clustering techniques are being explored in the implementation of such agents in computer games, these techniques are still under-used in the implementation of simulation systems. This work approaches this gap by proposing a new CBR and clustering framework in which clustering algorithms and clustering evaluation techniques are explored in both the construction of adjusted similarity functions and the organization of sub-case bases, which are indexing components to the efficient retrieval of relevant cases from case bases so as to support the solution of new simulation problems. To evaluate this framework, a case-based algorithm was implemented to simulate the choice of military supplies to be used in artillery battery missions in virtual tactical simulations.

Introduction

Military organizations rely on education and training to prepare individuals to act in difficult situations with a high level of efficiency. In this context, Artificial Intelligence (AI) techniques are been increasingly explored in the development of intelligent agent functionalities aiming to accurately reproduce real-world problem-solving situations in simulation environments. However, more dynamic and adaptive knowledge-based techniques still need to be explored to support the modeling and simulation of intelligent agent tasks in computer simulation systems. As investigated in this paper, these new kinds of AI approaches can emerge from the integration of Case-Based Reasoning (CBR) (Lopez De Mantaras, McSherry et al. 2005) and Clustering (Jain, N. et al. 1999) techniques. In an approach where CBR and clustering are integrated, clustering algorithms are used in the grouping of cases to investigate how well such cases capture different groups or classes representing real-world simulation situations. Most importantly, clustering is also used in an exploratory analysis involving such problem-solving experiences captured as cases, where clustering results by themselves have a paramount educational value for domain users which ought to be involved in the development of simulation systems.

This paper discusses a new CBR and clustering framework to support the development of case-based agents in virtual tactical simulation environments. For the enhancement of such systems, this work shows how to systematically use clustering algorithms in the solution of indexing problems in CBR in order to support the query and organization of case bases. The proposed framework innovates when it explores clustering algorithms and clustering evaluation metrics in the proposition and testing of domain-specific adjusted similarity functions to improve the accuracy of CBR queries in simulation systems. To evaluate the proposed CBR and clustering framework in a real-world simulation problem, a case study developed in this project involved the construction of a case-based agent algorithm for supporting the choice of type and amount of military supplies to be used by an artillery battery agent in a virtual tactical simulator (e.g. (Brondani, Freitas et al. 2017)).

The paper is organized as follows: the second section reviews basic concepts and related works from AI and Simulation, in the context of CBR and Clustering. The proposed CBR and Clustering framework along with experimental results are discussed in sections third and fourth, respectively. Finally, conclusions and future work are presented.

Background to this Work

Many agent development challenges are common to both simulation systems and computer game systems. For instance, planning and different degrees of prediction to tasks in a game, which are common issues for CBR in Real Time Strategy (RTS) games (Ontañón 2010, Lara-Cabrera, Cotta et al. 2013, Robertson and Watson 2014), are features that are also desirable in agents for simulation systems. In contrast to virtual simulations, where users are in-the-loop of the case reuse process (Reece, McCormack et al. 2004), CBR is most commonly used in fully autonomous agents as shown in an air combat simulator (Borck, Karneeb et al. 2015). Indexing structures to support CBR retrieval algorithms (Lopez De Mantaras, McSherry et al. 2005) are central to the success of such systems. To the identification of a set of weight values to be used in a similarity function of a case-based agent implementation, this work relies on clustering techniques. Clustering (Jain, N. et al. 1999) permits the investigation of how to organize cases into groups using similarities between them. Among standard approaches, the hierarchical algorithms allow the investigation of different levels of hierarchy associated to the formed groups, allowing clustering results to be naturally represented through tree structures. These results can be systematically evaluated by external metrics: "Entropy", "purity" and "precision" (Jain, N. et al. 1999). Different works demonstrated the usage of clustering techniques in the development of indexing approaches for case retrieval. Most of them are focused on the analysis of how large case bases could be organized in sub-case bases (e.g. (Yang and Wu 2000, Mittal, Sharma et al. 2014, Müller and Bergmann 2014)).

A CBR and Clustering Framework

The realism of experience-based agents can be achieved when CBR retrieval functionalities provide relevant similarity computations between a current simulation situation and past simulation problems recorded in a case base. In the modeling of agents for simulation systems, useful case-based solutions can be obtained provided that a pertinent set of attributes of a simulation problem is indexed in the case structure. In many systems, the relative importance of each attribute is represented by weight values in domain-specific weighted similarity computations. After weight values for these case attributes are set up in similarity functions, clustering algorithms relying on these adjusted functions can be executed multiple times.

When running most clustering algorithms, parameters need to be initially informed by users. Using these parameters, this framework describes how to explore clustering techniques in the construction of domain-specific similarity functions to be used by agents in simulation systems. As a result, CBR-based agents can use two-step retrieval algorithms, where the first step computes similarities between a query case against cluster centroids representing case groups formed when clustering algorithms are executed. Then, the second step computes similarities between the query case and the past cases that belong to the cluster centroid that is the most similar to the query as computed in the first step of retrieval. In the end, the most similar past cases belonging to its most similar sub-case base are retrieved from case bases, and the solutions recorded.

The indexing process regarding the assessment of the relevance of case attributes starts when **an equal weighting scheme** is explored in the underlying similarity function used in both clustering and CBR. Using this configuration of weights, this initial indexing activity is similar to what happens when standard clustering analysis tasks are executed. In this context, clustering results obtained in this initial investigation can be evaluated according to clustering evaluation methods. Using equally weighted similarity function in the clustering of cases, clustering algorithms are executed, and groupings of cases are recorded for domain user' inspection. Clustering results are also evaluated according to external clustering evaluation metrics (e.g. Table 1 (a)). In this described indexing process, it is possible to obtain "baseline" clustering evaluations that may be improved in the analysis of the indices.

Once baseline clustering results and clustering evaluations are recorded, the second indexing activity refers to the usage of **a single weighting scheme** in the similarity function used by clustering algorithms. The main idea is to select one case attribute used in the similarity computations and to assign a high weight value to this attribute, while all other weights for case attributes used in these similarity computations are kept equal to 1.0. Then, this single weighting scheme analysis considers that clustering algorithms are executed multiple times to consider each attribute represented in the case structure (e.g. Table 1 (b)). In doing so, the aim is to obtain clustering results along with clustering evaluations so that the impact of each case attribute in the formation of more homogeneous case groupings is assessed (e.g. Table 1 (b1), (b10) and (b11)). The impact in clustering results due to the use of high weight values in each case attribute used in similarity computations is also contrasted with similar evaluations made when the equal weighting scheme is evaluated (i.e. results are contrasted against values that are taken as "baseline" where equal weights for case attributes are used (e.g. Table 1 (a))). In practice, when the use of a high weight value in a target case attribute leads to the formation of better evaluated clustering results, this indicates that such targeted attribute has a higher relevance in the assessed similarity.

The last activity deals with **an adjusted weighting scheme** in similarity computations executed by clustering algorithms. To do so, the proposal here is that values of clustering evaluation obtained when the single weighting scheme is used (e.g. in Table 1 (b1), (b10) and (b11)) are automatically normalized into a set of attribute weight values. Using a linear normalization method, a range of values from 1.0 to 10.0 can easily allow domain experts to examine the resulting set of weights to assess whether they match such experts' knowledge with respect to the relative importance of case attributes in the solution of problems in their applications. Values of clustering evaluation obtained when the equal weighting scheme is executed ("baseline" clustering evaluations (e.g. Table 1 (a))) can also be explored in this normalization task. It means that clustering evaluation estimates from the single weighting scheme (i.e. clustering evaluations reflecting the impact of each case attribute in the formation of case groups) can be higher (or lower) than these baseline evaluations. In effect, when baseline clustering evaluations are used in this normalization, only the evaluation estimates that are higher than the baseline ones are normalized as a set of attribute weights, while the weights for other case attributes are kept to 1.0. Then, an adjusted set of weight values for case attributes is defined and used in domain-specific adjusted similarity functions for new clustering executions. It means that clustering algorithms are executed and clustering results obtained with their centroids and external evaluations (e.g. Table 1 (c)) are again recorded.

A Case for CBR and Clustering in the Development of Artillery Battery Agents

An evaluation of the CBR and clustering framework in the context of the development of (semi) autonomous agent behaviors for simulation systems was developed in a project involving the development of a virtual tactical simulator (e.g. (Brondani, Freitas et al. 2017)). In this project, the addressed problem regards the choice of supplies to be used by artillery battery agents to accomplish a mission in simulation exercises. Besides making careful considerations of available types and amounts of provisions, the solution to this problem involves the domain-specific assessment of contextual battle situation factors.

Experiments were carried out using a case base containing 800 cases for the artillery battery supply selection problem. This case base was organized to capture real-world decision-making situations, ranging from scenarios in which supplies were broadly available to scenarios in which such supplies were scarce and even not fully adequate to fulfill the mission goals. In these experiments, parameters for a hierarchical clustering algorithm used were configured as: number of cluster = 15 (empirically defined as they captured the main groups of cases), linking criteria used = "single", "complete" and "average". Table 1 presents clustering evaluation results from executions of a hierarchical clustering algorithm. In such various executions of this algorithm, the "equal weighting scheme", the "single weighting scheme" and the "adjusted weighting scheme" were explored in the similarity function used.

Table 1. Clustering evaluation results obtained when different configurations of the similarity function were used

Clustering algorithm	Linking strategy	Clustering evaluation metrics	Equal weighting scheme	Single weighting scheme					Adjusted weighting scheme	
			(a) All weight values for case attributes are equal to 1.0 (baseline clustering evaluations)	(b) Weight value for a selected case attribute = HIGH_VALUE, while all other weight values for case attributes = 1.0					(c) Weight values derived from the *linear normalization* of clustering evaluations	
				(b1) High weight attrib. 1	…	(b10) High weight attrib. 10	(b11) High weight attrib. 11	…	(c1) Baseline clustering evaluations are *not* used	(c2) Baseline clustering evaluations are used
Hierarchical clustering	Complete	Purity	0.08	0.08	…	0.33	0.14	…	0.08	0.09
		Entropy	-20.38	-25.89	…	-217.03	-43.05	…	-19.43	-20.79
		Precision	0.61	0.80	…	0.92	0.91	…	0.59	0.63
	Average	Purity	0.15	0.14	…	0.17	0.18	…	0.15	0.17
		Entropy	-35	-99	…	-145	-93	…	-34	-48
		Precision	0.53	0.62	…	0.63	0.58	…	0.52	0.64
	Single	Purity	0.07	0.09	…	0.23	0.21	…	0.08	0.13
		Entropy	-33	-52	…	-180	-165	…	-32	-40
		Precision	0.52	0.47	…	0.61	0.68	…	0.48	0.54

Table 2. Leave-one-out cross-validation results obtained when different configurations of the similarity function were used

Clustering algorithm	Linking strategy	Clustering evaluation metrics	(a) Accuracy results obtained when an equal weighting scheme is used		(b) Accuracy results (weight values derived from the *linear normalization* of clustering evaluations)			
					(b1) Baseline clustering evaluations are *not* used in the normalization task		(b2) Baseline clustering evaluations are used in the normalization task	
			(a1) Sub-case bases are *not* used	(a2) Sub-case bases are used	(b1.1) Sub-case bases are *not* used	(b1.2) Sub-case bases are used	(b2.1) Sub-case bases are *not* used	(b2.2) Sub-case bases are used
Hierarchical clustering	Complete	Purity	44.50	69.26	63.11	70.37	67.63	74.36
		Entropy	44.50	69.26	65.14	68.50	65.26	82.79
		Precision	44.50	69.26	58.38	65.40	68.00	79.88
	Average	Purity	44.50	60.72	57.52	62.00	61.65	69.23
		Entropy	44.50	60.72	61.66	67.54	64.14	73.42
		Precision	44.50	60.72	53.50	70.12	63.23	79.65
	Single	Purity	44.50	53.88	53.25	64.09	60.08	73.51
		Entropy	44.50	53.88	56.18	69.13	58.96	79.63
		Precision	44.50	53.88	57.05	61.22	63.40	71.62

Using the 800 cases available, a leave-one-out and test cross-validation method was used to assess the accuracy of the case-based agents implemented. Table 2 shows the accuracy results obtained when the equal weighting scheme was explored in the similarity function (Table 2 (a)). These results were divided in two sets of values to show if it is worth using sub-case bases (Table 2 (a2)) in the CBR retrieval algorithm. In this case, when sub-case bases were not used (Table 2 (a1)), clustering results were not used by the CBR algorithm since similarity computations were executed against all cases from the case base. When sub-case bases were used by the CBR retrieval algorithm, however, different linking strategies (complete, average and single) resulted in distinct organizations of case groups. In these tests, for instance, the CBR accuracy results obtained when the complete linking strategy was used in the hierarchical clustering (69.26) was higher than the accuracy results obtained (53.88 and 60.72) when the single and average strategies were used in the clustering (Table 2 (a2)). Table 2 (b) also shows the accuracy results obtained when different weighting schemes were used in the similarity function. These results were also divided in two sets of values to assess the impact of using baseline clustering evaluation values in the normalization task (Table 2 (b1) and (b2)). The accuracy results were again divided to show the impact of using sub-case bases in the retrieval of cases. When baseline clustering evaluations were not used in the normalization process leading to attribute weight values (Table 2 (b1)), the use of sub-case bases presented better accuracy results than when sub-case bases were not used (Table 2 (b1.1) compared to (b1.2)). The use of sub-case bases also provided relevant accuracy results when the normalization of attribute weights considered the baseline clustering evaluation values (Table 2 (b2.1) compared to (b2.2)). In the end, the best results were obtained when both the normalization of attribute weights considered the baseline clustering evaluation values and the sub-case bases produced from such adjusted weighting scheme were used in the retrieval of cases (Table 2 (b2.2)). In the implementation of CBR based agents in our virtual tactical simulation system, it was used the adjusted similarity function and the case base organization that produced the best accuracy result (82.79, which is well over the best baseline results: 44.50 and 69.26). This accuracy was derived when the "complete" linking strategy was used in the hierarchical clustering algorithm. This algorithm resulted in sub-case bases (15 case groups) used by a two step CBR retrieval algorithm, where the weights for the case attributes used in similarity computations were obtained from the normalization of "entropy" estimates.

Concluding Remarks

This paper addresses the integration of CBR and clustering techniques in the development of intelligent agent behaviors in realistic simulation environments. The main contribution of this framework is to show how to construct, evaluate and refine clustering results to support the determination of indexing structure for CBR systems. The usefulness of the proposed framework has been proven by CBR and clustering experiments developed for the supply selection agent algorithm that supports virtual tactical simulations of artillery battery tasks used in military training. Future work will be directed to the extension of the framework so that one could assess the combined use of other kinds of clustering techniques in the indexing of cases.

Acknowledgments

We thank the Brazilian Army for the financial support through the SIS-ASTROS Project (813782/2014), developed in the context of the PEE ASTROS 2020.

References

Borck, H., J. Karneeb, R. Alford and D. Aha. 2015. Case-based behavior recognition in beyond visual range air combat. *The 28th Int. Florida Artificial Intelligence Research Society Conf. (FLAIRS 2015)*. Florida, USA, AAAI Press**:** 379-384.

Brondani, J. R., E. P. Freitas and L. A. L. Silva. 2017. A task-oriented and parameterized (semi) autonomous navigation framework for the development of simulation systems. *The 28st Int. Conf. on Knowledge Based and Intelligent Information and Engineering Systems (KES 2017)*. Marseille, France, Procedia Computer Science. 112**:** 534–543.

Jain, A. K., M. M. N. and F. P. J. 1999. Data Clustering: A Review. *ACM Computing Surveys* 31(3): 264-323.

Lara-Cabrera, R., C. Cotta and A. J. Fernández-Leiva. 2013. A review of computational intelligence in RTS games. *IEEE Symposium on Foundations of Computational Intelligence (FOCI 2013)*. Singapore, IEEE**:** 114-121.

Lopez De Mantaras, R., D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T. Cox and K. Forbus. 2005. Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review* 20(03): 215-240.

Mittal, A., K. K. Sharma and S. Dalal. 2014. Applying clustering algorithm in case retrieval phase of the case-based reasoning. *Apurva Mittal al. Int. Journal of Research Aspects of Engineering and Management* 1(2): 14-16.

Müller, G. and R. Bergmann. 2014. A cluster-based approach to improve similarity-based retrieval for process-oriented case-based reasoning. *The 20th European Conf. on Artificial Intelligence (ECAI 2014)*. Prague, Czech Republic, IOS Press**:** 639-644.

Ontañón, S. 2010. On-line case-based planning. *Computational Intelligence* 26(1): 84-119.

Reece, D., J. McCormack and J. Zhang. 2004. A case-based reasoning tool for composing behaviors for computer generated forces. *Behavior Representation in Modeling and Simulation (BRIMS 2004)*. Arlington, VA.

Robertson, G. and I. Watson. 2014. A Review of Real-Time Strategy Game AI. *AI Magazine* 35(4).

Yang, Q. and J. Wu. 2000. Keep It Simple: A Case-Base Maintenance Policy Based on Clustering and Information Theory. *The 13th Biennial Conf. of the Canadian Society for Computational Studies of Intelligence*. Montreal, Canada, Lecture Notes in Computer Science. 1822**:** 102-114.