Case-based reasoning using expert systems to determine electricity reduction in residential buildings

Ricardo Faia[1], Tiago Pinto[1,2], Zita Vale[1], Juan Manuel Corchado[2]
[1]GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development
Polytechnic of Porto (ISEP/IPP), Porto, Portugal
{rfmfa, tmcfp, zav}@isep.ipp.pt
[2]BISITE Research Centre, University of Salamanca (USAL), Calle Espejo, 12, 37007 Salamanca, Spain
{tpinto, corchado}@usal.es

***Abstract*— Case-based reasoning enables solving new problems using past experience, by reusing solutions for past problems. The simplicity of this technique has made it very popular in several domains. However, the use of this type of approach to support decisions in the power and energy domain is still rather unexplored, especially regarding the flexibility of consumption in buildings in response to recent environmental concerns and consequent governmental policies that envisage the increase of energy efficiency. In order to determine the amount of consumption reduction that should be applied in a building, this article proposes a methodology that adapts the past results of similar cases in order to achieve a decision for the new case. A clustering methodology is used to identify the most similar previous cases, and an expert system is developed to refine the final solution after the combination of the similar cases results. The proposed CBR methodology is evaluated using a set of real data from a residential building. Results prove the advantages of the proposed methodology, demonstrating its applicability to enhance house energy management systems by determining the amount of reduction that should be applied in each moment, thus allowing such systems to carry out the reduction through the different loads of the building**

***Index Terms*— Case based reasoning, clustering, demand response, energy efficiency, expert systems, residential energy management**

I. INTRODUCTION

The study and development of Case-Based Reasoning (CBR) has been inspired by cognitive science research on human memory. In CBR the resolution of problems is based on memory, thus reflecting human use of remembered problems and solutions as a starting point for new problem solving. New observations on each problem solving are used in CBR, assuming that similar problems have similar solutions [1].

In CBR process, the new problems are solved by recognizing their similarity to a known problem and by adapting solutions that were used to solve the previous problems. The case indexing and retrieving processes are highly recommended, because the indexing and retrieving process of a similar cases in the database is important for modeling a successful CBR system. A CBR system that can reflect the experience of experts in building a system and solving a new problem is needed [2].

Cases are a contextualized piece of knowledge representing an experience. Many cases are gathered and stored in memory, in order to build a case base in a specific domain. This way, case base is composed of two spaces [3]. The new case of a problem is compared to the other cases stored in the case base and the most similar ones are extracted along with their associated solution. Then, the solution is adapted to reach a suitable solution to the new problem. When the new problem is solved, it is stored (or not) in the case base, depending on the retain policy.

The application of CBR has been highly successful in different areas, such as software estimation [4], document retrieval system development [5], mechanism failure identification [6], prediction of construction costs [7], bankruptcy prediction [8] and building new recipes [9], among many other.

Another relevant field of application is power and energy systems, although the number of works in this domain is not too significant. In [10] a CBR system for building energy prediction is proposed, with the aim at identifying operation issues and proposing better operating strategies. Simplified models based on CBR to predict the hourly electricity consumption of an institutional building are proposed in [11]. A CBR approach to predict electromagnetic flows is presented in [12]. A CBR method providing online decision-making for optimization of coal-blend combustion was investigated in [13]. The estimation of the energy performance of new buildings using CBR is studied in [14]. These are relevant contributions that cover some problems in the energy domain. However, many urgently needed issues in this area are still not addressed, such as considering consumers' flexibility.

The EU guidelines towards a low carbon society [15] and the recent EU Winter Package [16] frame consumers as a central piece in future power systems, being the consumption flexibility the most promising solution for the new challenges [17]. Building energy management systems play a crucial role in this scope, as they are able to support consumers' decisions

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 641794 (project DREAM-GO) and a grant agreement No 703689 (project ADAPT); and from FEDER Funds through COMPETE program and from National Funds through FCT under the project UID/EEA/00760/2013

in adapting their consumption without compromising their comfort [18]. However, deciding how much flexibility should be asked from each consumer when required is not an easy task as it depends on each individual consumer's habits and comfort.

This paper addresses the problem of deciding the amount of reduction that should be asked from consumers in moments when such is needed from the system. This is done by proposing a CBR system that uses previous cases of energy reduction in a building to decide which amount should be applied to a new case. A clustering approach is used to identify the most similar previous cases [19]. After these are selected, they are combined, and the final result is adapted using an expert system composed by several domain specific rules that refine the final solution.

After this introductory section, section 2 presents the proposed approach, with focus on the techniques that have been specifically developed for addressing this problem, namely the clustering approach to select the most similar cases, and the expert system to enable the adaptation of the final solution. Section 3 presents a case study using real data from a residential house, and section 4 presents the achieved results. Finally, section 5 presents the most relevant conclusions from this work.

II. System Architecture

This section presents the proposed methodology. The proposed system is based on the general CBR cycle. When a new case is presented, the most similar cases are retrieved, followed by the reuse task with which an adaptation of the solutions of the similar cases will be made for the solution of the new case. Afterwards, the revise phase uses an expert system to adapt and refine the final solution before the final retain phase. Fig. 1 shows the flowchart of the proposed methodology presented in this paper.

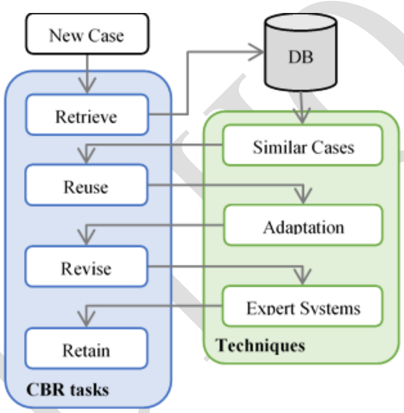

Fig. 1. CBR flowchart

As can be seen from Fig. 1 initially a new case is presented, for which it is necessary to find a solution. The CBR cycle starts with the retrieve task, where techniques are applied to select the most similar cases present in the DB. After this stage, the most similar cases are sent to the reuse task, which will use the similar solutions to reach a new solution, using adaptation techniques. Then, the task of reviewing is applied, where the created solution is reviewed. In this task, rules are created based on expert systems, in order to adapt and refine the final solution. In the end, the solution is evaluated so that the decision is made on whether to add the new case, or not, to the case base. The techniques used in the methodology are explained as follows.

A. *Similar cases*

The retrieve phase concerns the identification and retrieval of similar cases from the historic case-base log. This is done by using a clustering algorithm, namely the K-Means [20], to group the cases according to their similarity, thus identifying those that are most similar to the current case.

The clustering methodology considers a set of cases ($x_1, x_2, \ldots, x_n$), where each case is a d-dimensional real vector, and n is the number of considered cases. The clustering process aims at partitioning the n cases into k ($\leq n$) clusters $C = \{C_1, C_2, \ldots, C_k\}$ so that the Within-Cluster Sum of Squares (WCSS) is minimized (1).

$$\min \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \tag{1}$$

where μ_i is the mean of points in C_i, *i.e.* the cluster *centroid*.

The dimension of the vector that characterizes each case x_p, $p \in \{1, \ldots, n\}$ is equal to the sum of the individual dimensions of n vectors, where each of these n vectors contains the data that is referent to a different data variable, e.g. consumption, generation, tariff price, etc.

With the objective of minimizing equation (1), the clustering process executes an iterative process between two steps: (i) the assignment step, where each observation x_p is assigned to the cluster $C^{(t)}$ whose mean value yields the minimum WCSS in iteration t, as presented in (2); and (ii) the update step, where the new means of each cluster are calculated, considering the newly assigned cases, determining the new *centroid* μ_i of each cluster, as in (3).

$$C_i^{(t)} = \{x_p : ||x_p - \mu_i^{(t)}||^2 \leq ||x_p - \mu_j^{(t)}||^2 \; \forall j, 1 \leq j \leq k\} \tag{2}$$

$$\mu_i^{(t+1)} = \frac{1}{|C_i^{(t)}|} \sum_{x_j \in C_i^{(t)}} x_j \tag{3}$$

The execution of the algorithm stops when the convergence process is completed, *i.e.* when the assignments of cases to different clusters no longer change. By minimizing the WCSS objective, in equation (1), the K-Means clustering methodology assigns cases to the nearest cluster by distance. This means that each case will be grouped in the same cluster as the ones that are more similar. These are the cases that are retrieved by the CBR process.

B. *Adaptation*

In the adaptation is obtained the result value for the new case, considering the result that each similar case that has been identified in the retrieve phase. At the end of this step three different results can be obtained, which are the average of the results of the similar cases, the maximum and the minimum value. Equation (4) is applied to obtain the mean value, where CR is a result to this problem.

$$CR_{result}^{mean} = \frac{\sum CR_{result}^{similar}}{number\ of\ similar\ cases} \tag{4}$$

Equation (5), is applied to obtain the maximum value.

$$CR_{result}^{maximum} = \max\left(CR_{result}^{similar}\right) \tag{5}$$

Equation (6), is applied to obtain the minimum value.

$$CR_{result}^{minimum} = \min\left(CR_{result}^{similar}\right) \tag{6}$$

C. Expert systems

Considering the existing knowledge of the problem, the revise task is applied using an expert system. Expert systems intend to emulate in CBR systems the decision-making ability of a human expert. In this case rules are created, which are applied to the value obtained in the reuse process. Most of the created rules take into account the time of day: x_3, the description about other variables is present in Table 1. The objective of this expert system is to adapt the solution when it is not coherent, e.g. if it is asked for the consumer to reduce more than its total expected consumption. In equation (7), if the result from the reuse task is smaller than zero, then the result of the revision is 0.

$$if\ CRreuse < 0\ \rightarrow CRrevise = 0 \tag{7}$$

In equation (8) the $CRreuse$ and $CRrevise$ represent the result of CBR in reuse task and revise, and is defined the rule for the hours between 0 and 5. In this rule is imposed on the system that at least there must be 0.8 kW of consumption in the end, and reduce the remaining by half.

$$f\ x_3 \geq 0\ \cap\ x_3 < 5\ \cap\ (CRreuse = 0\ \cup\ x_8 > 0.8)\ \rightarrow CRrevise = \frac{x_8 - 0.8}{2} \tag{8}$$

Equation (9) represents the mealtimes, and at this time the reduction should be 25% of the value corresponding to the load x_8 (represent the electric consumption) less the production itself x_9 (represent the electric generation).

$$\begin{aligned} if\ \big((\ x_3 > 7.3\ \cap\ x_3 < 9) \cup (\ x_3 > 12\ \cap\ x_3 < 13.3) \\ \cup (\ x_3 > 19.3\ \cap\ x_3 < 21)\big) \\ \cap (CRreuse = 0\ \cup CRreuse > x_8 - x_9) \\ \rightarrow CRrevise = 0.25 \times (x_8 - x_9) \end{aligned} \tag{9}$$

Equation (10) represents the hours between breakfast and lunch and from lunch to dinner, this rule takes into account the number of people in the housing, and if it is less than 3, the value of the reduction will be half of consumption minus the production, otherwise the reduction will be a quarter, x_3 represent the hour.

$$\begin{aligned} if\big((\ x_3 > 9\ \cap\ x_3 < 12) \cup (\ x_3 > 14\ \cap\ x_3 < 18)\big) \\ \cap (CRreuse = 0 \cup CRreuse > x_8 - x_9) \\ elseif\ x_7 < 3 \rightarrow CRrevise = \frac{x_8 - x_9}{2} \\ else\ CRrevise = \frac{x_8 - x_9}{4} \end{aligned} \tag{10}$$

Equation (11) represents the hours between the 21 and the 0 of the following day. The value of reduction will be half of the load less the production; however it is expected that the production in this schedule is of 0 kW or approximately, x_7 represent the persons number.

$$\begin{aligned} if\ x_3 > 21\ \cap\ x_3 < 0\ \cap (CRreuse = 0 \cup CRreuse > x_8 - x_9) \\ \rightarrow CRrevise = \frac{x_8 - x_9}{2} \end{aligned} \tag{11}$$

After applying the rules to the value obtained in the reuse process, the revised value is achieved, thus making it possible to turn incoherent results into applicable solutions.

III. Case Study

The proposed CBR approach requires a case-base with historic cases. The considered case-base is constructed from previous cases regarding a test building to which the model is applied. The case-base has 11 different variables which are collected and recorded from sensors and other type of data collection. Table 1 shows all the stored variables as well as the type of each variable. The case-base is filled with the historic cases with normalized values. The normalization process converts raw values to standart scores, this process requires seleting values that span one range and representing them in another range. Assumption that the data are aproximated by a normal distribution, this is converted to a standard normal distribution, where the mean is 0 and the standard deviation is 1, applying the normal distribution. Table 2 shows the cases to which the presented model is applied in this case study (i.e. the reference cases to which the results from the proposed model are compared).

Table 1. Configuration of case variables

Representation	Name of variables	Score values	Units
x_1	Weekday	$\{1\ to\ 7\}$	-
x_2	Month	$\{1\ to\ 12\}$	-
x_3	Hour	$\{1\ to\ 24\}$	h
x_4	Season	$\{1\ to\ 4\}$	-
x_5	External Temperature	$\{7{,}1\ to\ 29{,}1\}$	ºC
x_6	External Humidity	$\{46\ to\ 98\}$	%
x_7	Persons Number	$\{1\ to\ 4\}$	-
x_8	Electric consumption	$\{0{,}1\ to\ 3{,}693\}$	kW
x_9	Electric Generation	$\{0\ to\ 2\}$	kW
x_{10}	Electric Tariff	$\{0{,}1634\ \}$	€/kW
R	Electric Reduction	$\{0\ to\ 0{,}8\}$	kW

Table 2. Profile of case for validation

Case	Case 1	Case 2	Case 3	Case 4
x_1	7	1	7	7
x_2	1	8	10	10
x_3	2,45	9,3	18,45	23,45
x_4	3	1	2	2
x_5	11,2	17,5	20,2	20,6
x_6	95	86	89	67
x_7	4	4	4	6
x_8	0,6	0,12	3,435	2,413
x_9	0	0,98	0,125	0
x_{10}	0,1634	0,1634	0,1634	0,1634
R_{real}	0	0	0,05	0,09

The cases presented in Table 2 are taken from the case-base, and when the model is applied these results will not be present in the case-base of historic cases. The selection was made taking into account the time of the day, choosing different hours. The validation of the CBR model will be done through the comparison of the real values and the values proposed by the CBR. Each case of those shown in Table 2 is considered as a new case.

IV. Results

This section shows the results of the case study. In order to analyze the impact of the numbers of clusters, which determine the number of similar cases, three different options are implemented: 20 clusters, 10 clusters and 5 clusters. Table 3 shows the main results of the proposed methodology. The value of Root Mean Square Error (RMSE) is expressed in Table 3, and gives the indication about the CBR system precision in a prediction value. As it is possible to observe the value of RMSE undergoes changes with the number of clusters. In this case, it is not possible to relate the number of clusters with the RMSE, because as one can see, with 20 clusters the error is the smallest one following the error with 5 clusters and ending with the error of the 10 clusters as the worst one.

Table 3. Results of the proposed methodology

Cases		Case 1	Case 2	Case 3	Case 4	**RMSE**
20 clusters	$\boldsymbol{R_{real}}$	0	0	0,05	0,09	0,124
	$\boldsymbol{CR}$*	0,024	0,060	0,238	0,238	
10 clusters	$\boldsymbol{R_{real}}$	0	0	0,05	0,09	0,141
	$\boldsymbol{CR}$*	0,000	0,014	0,040	0,026	
5 clusters	$\boldsymbol{R_{real}}$	0	0	0,05	0,09	0,135
	$\boldsymbol{CR}$*	0,000	0,007	0,040	0,026	

*Mean values

One of the most important factors in this system is the execution time that is necessary for a solution to be obtained, because the configuration of the loads in the house can change in any instant and the value of reduction will not be the same. For this reason, it is necessary to obtain a result in a short time, Fig. 2 shows the average execution time per number of clusters.

As one can see by Fig. 2, the execution time is longer when the number of clusters is larger, and it takes about 8.3 seconds, which is still acceptable. The trend is for the execution time to decrease as the number of clusters decreases. This is explained by analyzing the data in Table 4, which expresses the time elapsed in each task.

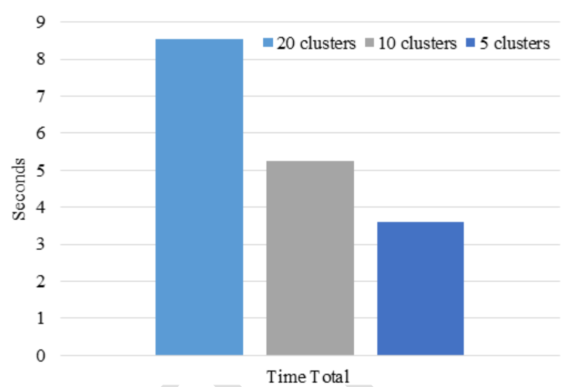

Fig. 2. Average execution time *per* number of clusters

Table 4. Time results for different tasks

Number of clusters		20	10	5
Time (seconds)	Load	1,372629	1,329359	1,30168
	Retrieve	7,152161	3,910887	2,307143
	Reuse	0,005667	0,005667	0,005667
	Revise	0,001921	0,000586	0,00056
	Retain	2,40E-06	1,12E-06	1,04E-06

From Table 4 it is possible to analyze the execution times by task, and it is possible to see that the task that takes more time is the retrieve, which is responsible for the selection of the most similar cases. In this task, the K-Means technique is implemented which, as we can see, will make a difference in the final result of the execution time. The higher the number of clusters, the longer this task takes to run. In Table 5 is shown the influence of the number of clusters in the selection of the most similar cases.

Table 5. Index of similar cases

Target Case	Nº clusters	Index of similar cases	Nº similar cases
1	20	$\{10, 12, 15, 48, 49, 55, 57\}$	7
	10	$\{1, 10, 12, 15, 48, 49, 55, 57\}$	8
	5	$\{1, 10, 12, 15, 17, 21, 31, 35, 48, 49, 55, 57\}$	12
2	20	$\{19\}$	1
	10	$\{19, 32, 56, 59\}$	4
	5	$\{3, 19, 24, 32, 56, 58, 59\}$	7
3	20	$\{9, 11, 13, 16, 27, 28, 29, 50, 53\}$	9
	10	$\{2, 5, 9, 11, 13, 16, 27, 28, 29\}$	9
	5	$\{2, 5, 9, 11, 13, 16, 27, 28, 29, 34, 37, 40, 50, 5$	14
4	20	$\{9, 11, 13, 16, 27, 28, 29, 50, 53\}$	9
	10	$\{2, 5, 9, 11, 13, 16, 27, 28, 29\}$	9
	5	$\{2, 5, 9, 11, 13, 16, 27, 28, 29, 34, 37, 40, 50, 5$	14

As can be seen from the Table 5, the number of similar cases tends to increase as the number of clusters is smaller. This is as expected, because if the number of clusters is smaller, more elements will contain each cluster, thus more elements will belong to the cluster in which the new case is inserted. Another interesting fact is the tendency of repetition of similar cases when using more or less clusters. This happens in general, as the number of similar cases increases (with less clusters), new cases are added and the existing cases are repeated.

Fig. 3 shows an illustration of case 2 in the selection of the most similar cases. In the three graphs, all cases of the database are expressed and in the y-axis the value of the result is presented. The presented data are normalized.

The different graphs of Fig. 3 show the similar cases selected in the different simulation for case 2. Initially in Fig. 4 a), with 20 clusters only one similar case is selected (point with blue circle), in Fig. 4 b), we have 10 clusters and 4 similar cases are selected; finally, in Fig. 4 c) 5 clusters are used and 7 similar cases selected. The number of selected cases is equal to the number of elements belonging to the cluster of the new case.

V. Conclusions

This paper presented a CBR approach that uses previous cases to determine the amount of energy reduction that should be applied in residential houses. A clustering methodology is used to identify similar cases, and an expert system to determine a set of rules used to adapt and refine the final solution.

From the achieved results, it can be concluded that the proposed CBR methodology is able to successfully solve the envisaged problem. As can be seen, the final result always depends on the number of clusters selected a priori, but by the result of the error can be verified that the variation is minimal. In the set of the three simulations it was obtained a 13.3% error for the validation set. At the end of the runtime, any of the simulations is applicable, since the time the load of the dwelling takes to vary will be greater than 8.3 seconds (maximum time taken in the simulations). The clustering technique worked correctly as it is possible to see by the results, because when the number of clusters is greater, fewer similar cases will be selected and when the number of clusters decreases more similar cases are selected.

As the future work, it is intended to try different techniques of case selection, as the case of *k*-nearest neighbor and decision trees. In the task of reuse, the use of weights in order to enable the weighing of the most similar cases is also a topic that can be addressed in the future.

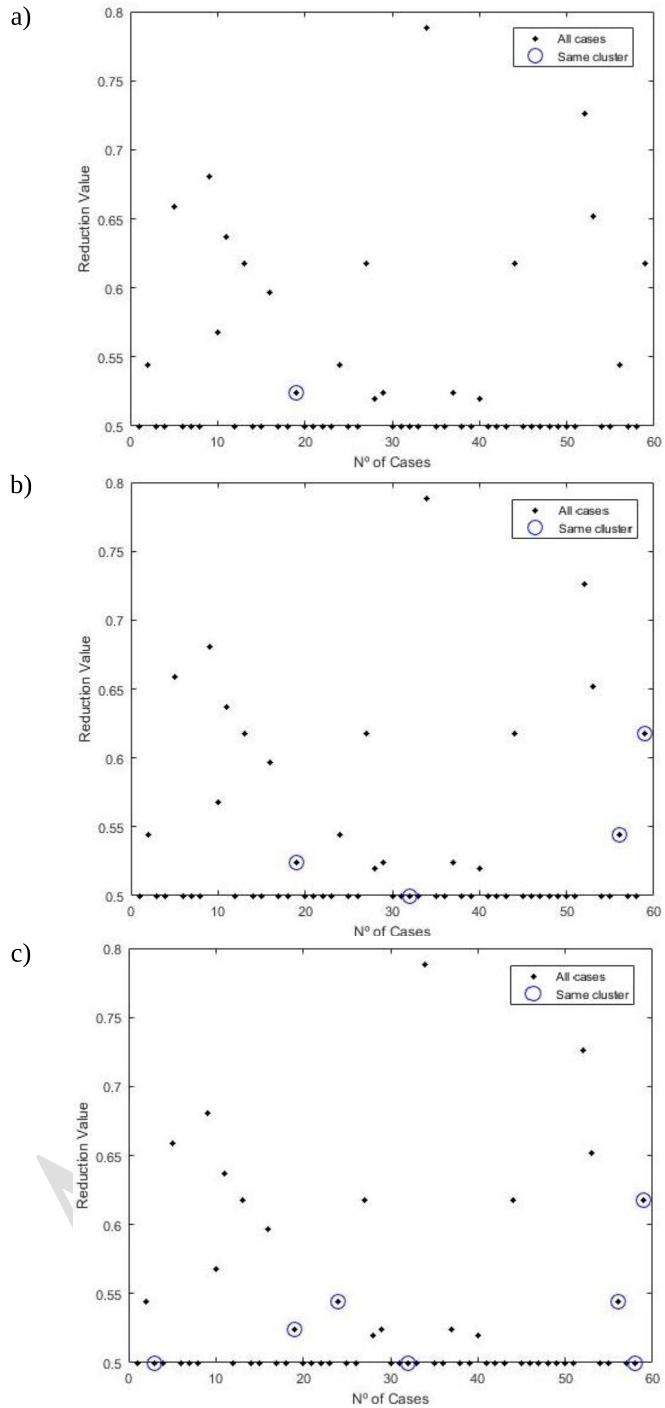

Fig. 3. Index of similar cases when using: a) 20 clusters, b) 10 clusters and c) 5 clusters

REFERENCES

[1] R. Lopez De Mantaras et al., "Retrieval, reuse, revision and retention in case-based reasoning," Knowl. Eng. Rev., vol. 20, no. 3, p. 215, Sep. 2005.

[2] S.-H. An, G.-H. Kim, and K.-I. Kang, "A case-based reasoning cost estimating model using experience by analytic hierarchy process," Build. Environ., vol. 42, no. 7, pp. 2573–2579, 2007.

[3] N. Stéphane and L. L. J. Marc, "Case-based reasoning for chemical engineering design," Chem. Eng. Res. Des., vol. 86, no. 6, pp. 648–658, 2008.

[4] G. R. Finnie, G. E. Wittig, and J.-M. Desharnais, "Estimating software development effort with case-based reasoning," in Case-Based Reasoning Research and Development: Second International Conference on Case-Based Reasoning, ICCBR-97 Providence, RI, USA, July 25--27, 1997

[5] I. Watson and H. Watson, "CAIRN: Case-Based Information Retrieval," in Proceedings of the 3rd United Kingdom Case-Based Reasoning Workshop, 1997.

[6] T. W. Liao, Z. M. Zhang, and C. R. Mount, "A case-based reasoning system for identifying failure mechanisms," Eng. Appl. Artif. Intell., vol. 13, no. 2, pp. 199–213, 2000.

[7] S.-Y. Kim, J.-W. Choi, G.-H. Kim, and K.-I. Kang, "Comparing Cost Prediction Methods for Apartment Housing Projects: CBR versus ANN," J. Asian Archit. Build. Eng., vol. 4, 2005.

[8] H. Ahn and K. Kim, "Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach," Appl. Soft Comput., vol. 9, no. 2, pp. 599–607, 2009.

[9] K. J. Hammond, "Case-Based Planning: Viewing Planning as a Memory Task," 1989.

[10] D. Monfet, M. Corsi, D. Choinière, and E. Arkhipova, "Development of an energy prediction tool for commercial buildings using case-based reasoning," Energy Build., vol. 81, pp. 152–160, Oct. 2014.

[11] R. Platon, V. R. Dehkordi, and J. Martel, "Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis," Energy Build., vol. 92, pp. 10–18, Apr. 2015.

[12] X. Huang and S. Yue, "The Energy-Saving Strategy of Electromagnetic Flowmeters Based on Various Predicting Methods," IEEE Transactions on Applied Superconductivity, vol. 26, no. 7. pp. 1–5, 2016.

[13] J. Xia, G. Chen, P. Tan, and C. Zhang, "An online case-based reasoning system for coal blends combustion optimization of thermal power plant," Int. J. Electr. Power Energy Syst., vol. 62, pp. 299–311, Nov. 2014.

[14] T. Hong, C. Koo, D. Kim, M. Lee, and J. Kim, "An estimation methodology for the dynamic operational rating of a new residential building using the advanced case-based reasoning and stochastic approaches," Appl. Energy, vol. 150, pp. 308–322, Jul. 2015.

[15] European Commission, "The 2020 climate and energy packages," Climate Action, 2009. [Online]. Available: http://ec.europa.eu/clima/policies/package/index_en.htm. [Accessed: 02-Oct-2015].

[16] European Commission, "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on ensuring the cross-border portability of online content services in the internal market," Brussels, 2015.

[17] B. Teixeira, F. Silva, T. Pinto, I. Praça, G. Santos, and Z. Vale, "Data mining approach to support the generation of Realistic Scenarios for multi-agent simulation of electricity markets," Intelligent Agents (IA), 2014 IEEE Symposium on. pp. 8–15, 2014.

[18] F. Fernandes, H. Morais, Z. Vale, and C. Ramos, "Dynamic load management in a smart home to participate in demand response events," Energy Build., vol. 82, pp. 592–606, 2014.

[19] R. Faia, T. Pinto, and Z. Vale, "Dynamic Fuzzy Estimation of Contracts Historic Information Using an Automatic Clustering Methodology," in Communications in Computer and Information Science, vol. 524, Springer International Publishing, 2015, pp. 270–282.

[20] J. Han, M. Kamber, and J. (Computer scientist) Pei, Data mining : concepts and techniques. Elsevier/Morgan Kaufmann, 2012.