Predicting the Electricity Consumption of Buildings: An Improved CBR Approach

Aulon Shabani[1], Adil Paul[2], Radu Platon[3], and Eyke Hüllermeier[2]

[1] Polytechnic University of Tirana
Tirana, Albania
aulon.shabani@fie.upt.al

[2] Department of Computer Science
Paderborn University, Germany
{adil.paul,eyke}@upb.de

[3] Natural Resources Canada
CanmetENERGY
Varennes, Canada
radu.platon@nrcan.gc.ca

Abstract. Case-based reasoning has recently been used to predict the hourly electricity consumption of institutional buildings. Past measurements of the building's operation are modeled as cases and, combined with forecast weather information, used to predict the electricity demand for the next six hours. Elaborating on this idea, we present an improved CBR approach that yields more accurate predictions of energy consumption. In particular, we develop improved (local) similarity measures specifically tailored for this kind of application, and combine these measures with a regression-based method for similarity learning. Moreover, we incorporate a simple procedure for case adaptation. Experimental results for a real case study confirm a significant improvement in predictive accuracy compared to previous approaches.

1 Introduction

Buildings are major energy users, being responsible for more than one-third of the world's total energy consumption [1]. In North America (U.S. and Canada) alone, institutional and commercial buildings account for 40% of total energy use [3]. A significant proportion of a building's energy consumption is used to operate increasingly complex systems and technologies, such as advanced mechanical heating, ventilation and air conditioning (HVAC) systems and thermal storage systems, designed to store energy for proper subsequent use.

Building operation and control need to be improved in order to reduce energy use, which becomes more and more a priority due to increasing energy prices and operation costs. The use of intelligent technologies enabling buildings to become proactive, by adapting their operation according to changing operational and environmental conditions can have a major impact on energy consumption. According to the Energy Star Program, energy consumption of commercial and institutional buildings can be reduced by up to 35% by using intelligent technologies and by modifying control practices [5].

Forecasting building energy use is critical for optimizing the management of thermal energy storage systems and for improving control and operation sequences in order to reduce energy consumption. It also enables energy use monitoring in order to identify periods of excessive consumption. Estimating the electricity consumption ahead of time enables improved planning of the operation of thermal energy storage devices linked to electrically-driven HVAC systems, optimizing their use and reducing peak loads and costs.

Different predictive models have been proposed for building energy use, mostly based on data-driven (machine learning) methods that require a significant amount of a building's historical operational data. However, data of that kind is not available for all buildings, such as in the case of new and retrofit buildings that underwent major changes to the point that previous data is no longer representative of current operation. As argued by Platon et al. [19], case-based reasoning offers a quite appealing alternative, not only due to being more transparent than black-box models like neural networks, but also due to its ability to operate with even little experience, and to learn and improve predictive accuracy as more data becomes available. Adding to this, we like to mention the potential of CBR to properly adapt predictions from previous to similar problems (such as retrofit buildings).

Recently, first promising results could indeed be achieved with a CBR model for predicting electricity use in an institutional facility over a time horizon of 6 hours [19, 20]. However, the predictive error of that model was still almost twice as high as that of a neural network, which severely hampers the willingness of building owners and operators to adopt this type of model: as decisions regarding building operation and control are made using the predicted energy consumption, the accuracy of the model is crucial for optimal operation and planning. Therefore, this paper presents various improvements made to the CBR model that led to a significant increase in predictive accuracy.

The rest of the paper is organized as follows. We start with a short overview of related work on energy prediction, prior to recalling the CBR model of [19, 20] in Section 3. Our improved approach in then presented in Section 4. In Section 5, this approach is empirically evaluated using data from an institutional Canadian facility located in Calgary, prior to concluding the paper in Section 6.

2 Predicting energy demand in buildings

Different types of methods for predicting energy demand in buildings have been proposed in the literature, including model-based approaches, statistical time series analysis, and machine learning methods.

Model-based approaches make use of a building's characteristics, such as total heating and cooling demand, thermal characteristics of walls, windows, other material proprieties, solar radiation, etc., in order to develop mathematical models for the simulation of the building's energy performance. Typical examples of such approaches are DOE-2, BLAST, EnergyPlus (a combination of DOE-2 and BLAST), SPARK, and TRNSYS; for a detailed description of the most commonly used simulation tools, we refer to [8].

The design of simulation models is a costly and time-consuming process, which requires a significant amount of expert knowledge. As an alternative, machine learning methods such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) can be used to induce models for energy demand prediction in a data-driven way, i.e., on the basis of energy demand observed in the past. For example, Azadeh et al. [4] train multi-layer perceptrons for predicting annual energy consumption of high energy consumers in the industrial sector. Likewise, Gonzalez and Zamarreno [12] predict energy consumption using a recurrent neural network. Using real data and taking forecast temperature values as attributes, highly precise results are achieved. Hybrid approaches combining simulation models with neural networks can be found in the literature, too, for example to predict energy consumption of a passive solar building [14]. Examples of prediction methods based on SVMs include [17] and [18]. A detailed review of machine learning methods for the prediction of a building's energy consumption is provided by [23].

As already mentioned in the introduction, CBR has been put forward as yet another alternative for the purpose of predicting a building's energy consumption more recently [16, 15, 19, 20]. Compared to standard (model-based) machine learning methods like ANN and SVM, case-based reasoning arguably comes with a number of advantages. In particular, since CBR is an inherently incremental process, it is able to adequately deal with an initial absence of historical consumption data, while continuously improving when more data becomes available over time. Moreover, CBR appears to be especially appealing for realizing knowledge transfer from one building to another, i.e., for exploiting data about one building to improve predictions for different yet similar buildings. First results on the use of CBR for energy prediction are promising and adhere to the limits recommended by the ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) [2]. More details about CBR for energy prediction are provided in the following section.

3 CBR for predicting electricity consumption

Since our work mainly builds on [19, 20], we devote this section to a short overview of these approaches, prior to presenting our improved method in Section 4. Platon et al. are interested in predicting hourly energy consumption based on historical measurements. To this end, they proceed from a case representation as shown in Table 2. Each case provides information about the development of 10 variables $V_1, \ldots, V_{10}$ (see Table 1) measured over 9 hours. The query case contains values of these variables for the current hour (t_0) as well as the previous two hours (t_{-1} and t_{-2}). Moreover, for the two variables air temperature and humidity, it contains predicted values over a period of 6 hours. The goal is to predict the electricity consumption over these 6 hours. The source case (memo-

Table 1: Variable description and measurement unit

	Variable	Unit
V_1	Forecast outside air temperature	(°C)
V_2	Forecast outside air relative humidity	(%)
V_3	Air handling unit 2 supply hot air temperature	(°C)
V_4	Air handling unit 3 supply hot air temperature	(°C)
V_5	West wing air handling unit supply cold air temperature	(°C)
V_6	Air handling unit 4 supply cold air temperature	(°C)
V_7	Chiller outlet water temperature	(°C)
V_8	Chiller outlet water flow rate	(l/s)
V_9	Boiler outlet water temperature	(°C)
V_{10}	Boiler outlet water flow rate	(l/s)

rized in the past) comprises the same information, though with real (instead of forecast) values for temperature and humidity; besides, the values for the target variable, electricity consumption, are given, too.

In the following, we denote by $x_{i,j}$ the value of the variable V_i at time point t_j in the source case ($1 \leq i \leq 10$, $-2 \leq j \leq 6$), and by p_j the value of the consumption P at time t_j. The corresponding values for the query case are denoted $y_{i,j}$ and q_j. The measurements of each variable V_i over time are collected in the time series x_i and y_i, respectively (corresponding to individual columns in the case representation). The combination of all values are referred to as X (source case) and Y (query case), respectively.

Similarity between cases is derived in two steps. First, given a new query case, only those previous cases are considered that fulfill the following properties: The time t_0 differs by at most one hour, and the absolute temperature at t_0 differs by at most $2°C$. Since the temperature and the time of the day are two very important properties, this can be seen as a prefiltering of presumably irrelevant cases (the similarity of which is formally set to 0).

For all other cases, the similarity is defined as a weighted average of the similarities of the different (input) variables:

$$\mathrm{CS}(X, Y) = \sum_{i=1}^{M} v_i \cdot \mathrm{VS}'_i(x_i, y_i) \ , \tag{1}$$

where $M = 10$ is the number of variables, $v_i \geq 0$ is the weight of the variable V_i, and $\mathrm{VS}'_i(x_i, y_i)$ the (local) similarity of the cases on that variable. As illustrated in Figure 1, variable similarity is defined as

$$\mathrm{VS}'_i(x_i, y_i) = \begin{cases} 0 & \text{if } D_w(x_i, y_i) > d^i_{max} \\ \frac{D_w(x_i,y_i) - d^i_{min}}{d^i_{max} - d^i_{min}} & \text{if } d^i_{min} \leq D_w(x_i, y_i) \leq d^i_{max} \\ 1 & \text{if } D_w(x_i, y_i) < d^i_{min} \end{cases} \ . \tag{2}$$

Table 2: Example of a query and a source case. Numbers in blue in the query case are forecast. Numbers in gray in the source case are known but not used for comparison with the source case (for which they are not given).

query case

	date and time	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	P
t_6	2014-04-07 15:00	10	32.2									?
t_5	2014-04-07 14:00	9	39.9									?
t_4	2014-04-07 13:00	8	32.2									?
t_3	2014-04-07 12:00	9	33.8									?
t_2	2014-04-07 11:00	10	34.6									?
t_1	2014-04-07 10:00	11	29.4									?
t_0	2014-04-07 09:00	12	29.9	29.4	28.4	15.8	24.9	30.5	-.05	67.3	76.2	203.1
t_{-1}	2014-04-07 08:00	12	31.2	29.6	17.4	10.1	21.4	32.7	-.05	65.2	76.2	203.8
t_{-2}	2014-04-07 07:00	11	31.0	28.3	9.7	10.6	22.4	30.8	-.04	66.3	73.2	197.6

source case

	date and time	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	P
t_6	2014-03-06 14:00	12	29.9	25.4	24.7	17.9	22.2	31.2	-.04	63.6	77.1	202.9
t_5	2014-03-06 13:00	12	29.9	28.4	27.4	16.7	25.2	31.2	-.04	65.8	78.1	204.6
t_4	2014-03-06 12:00	12	29.9	31.1	29.1	17.4	26.1	31.4	-.05	69.1	77.1	205.1
t_3	2014-03-06 11:00	11	20.5	27.4	27.4	19.8	22.6	29.4	-.04	71.1	76.8	202.1
t_2	2014-03-06 10:00	12	28.4	29.2	28.8	15.9	24.2	30.6	-.05	67.8	76.0	203.8
t_1	2014-03-06 09:00	11	29.3	27.4	25.4	17.8	26.9	31.5	-.05	66.3	77.2	204.6
t_0	2014-03-06 08:00	12	29.4	29.2	28.4	16.8	22.9	31.5	-.05	69.3	77.2	204.2
t_{-1}	2014-03-06 07:00	11	31.2	29.8	17.6	10.1	22.4	32.7	-.05	65.2	76.2	204.8
t_{-2}	2014-03-06 06:00	10	31.0	28.1	9.8	11.6	21.4	31.8	-.04	65.3	73.2	199.8

Here, d^i_{min} and d^i_{max} are variable-specific thresholds specifying what can be seen as completely similar and completely dissimilar (cf. Table 3), and D_w is the weighted Euclidean distance:

$$D_w(x_i, y_i) = \frac{\sqrt{\sum_{j=-2}^{n} w_j (x_{i,j} - y_{i,j})^2}}{\sum_{j=-2}^{n} w_j} \quad , \tag{3}$$

where $n = 0$ or $n = 6$ (depending on the variable), and the weights $w_j = 1 + j/3$ for $j \in \{-2, -1, 0\}$ and $w_j = 1 - j/7$ for $j \in \{1, \dots, 6\}$ are such that observations closer to the current time t_0 have a higher influence.

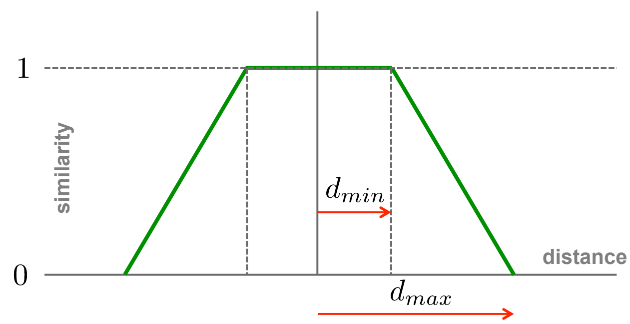

Fig. 1: Transformation of Euclidean distance into similarity.

At prediction time, given a query case Y, those previous cases $X_1, \dots, X_K$ with similarity $\text{CS}(X_k, Y) > 0.8$ are retrieved from the case base, and predictions of energy consumption are obtained as weighted averages of the consumptions observed for these cases:

$$\hat{q}_j = \frac{\sum_{k=1}^{K} \text{CS}(X_k, Y) \cdot p_{k,j}}{\sum_{k=1}^{K} \text{CS}(X_k, Y)} \quad , \tag{4}$$

where $p_{k,j}$ is the consumption for case X_k at time $j \in \{1, \dots, 6\}$.

As commonly done in the electricity and energy domain[4], predictive performance is measured in terms of the CV-RMSE (Coefficient of Variation Root Mean Square Error): With $\{\hat{q}_t \,|\, t \in T\}$ a set of predicted consumptions (for a single but possibly also for several query cases) and $\{q_t \,|\, t \in T\}$ the corresponding observed values, this measure is defined as

$$\text{CV-RMSE} = \frac{\sqrt{\frac{1}{|T|-1} \sum_{t=1}^{|T|} (q_t - \hat{q}_t)^2}}{\overline{q}} \times 100 \quad , \tag{5}$$

where $\overline{q}$ is the mean of true values.

[4] ANSI/BPI-2400-S-2012 Standard Practice for Standardized Qualification of Whole-House Energy Savings Predictions by Calibration to Energy Use History

4 Improved CBR model

Building on the CBR model as outlined in the previous section, we devised a number of improvements that will be described in the following.

4.1 Variable similarity

According to (2), the similarity between two measurement sequences on a variable is a non-linear transformation of the Euclidean distance between these two sequences. While Euclidean distance is an established and reasonable measure, it arguably fails to properly account for the *trend* in the corresponding time series. Needless to say, looking at the trend is important when it comes to extrapolating into the future. For example, Figure 2 shows the time series for a specific variable (amplitude) and three cases. According to Euclidean distance, the first one (green, solid line) is as similar to the second (blue, short dashes) as to the third one (orange, long dashes). Looking at the trend, however, the third one appears to be much more relevant. In particular, the third case seems to be much more amenable to adaptation (cf. Section 4.3 below).

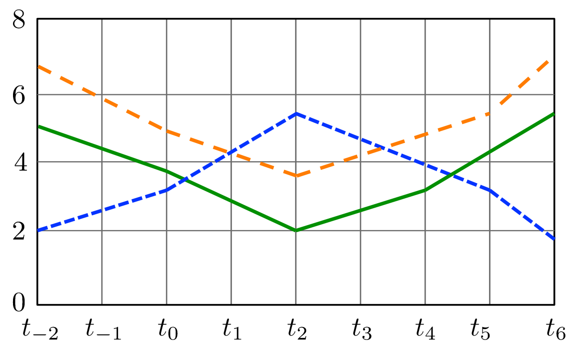

Fig. 2: Example of time series with different shape.

To capture the trend of time series, we define a second (variable) similarity measure based on the well-known cosine similarity [10, 9]: A sequence of values $x_i = (x_{i,-2}, x_{i,-1}, \ldots, x_{i,n})$ is considered as a bundle of two-dimensional vectors[5]

$$\left\{ (\Delta x_j, \Delta t_j)^\top = (x_{i,j+1} - x_{i,j}, 1)^\top \,|\, j = -2, \ldots, n-1 \right\} ,$$

and similarity is defined as the averaged (normalized) angle between the corresponding vectors (cf. Figure 3):

$$\mathrm{VS}_i''(x_i, y_i) = \frac{1}{\pi(n+2)} \sum_{j=-2}^{n-1} \cos^{-1}\left(\frac{\Delta x_j \Delta y_j + 1}{\sqrt{(\Delta x_j)^2 + 1}\sqrt{(\Delta y_j)^2 + 1}} \right) \qquad (6)$$

[5] In our case, the difference between time steps, Δt_j, is always 1, because measurements are made on an hourly basis.

Finally, we define a new variable similarity measure in terms of a weighted average of the original measure (2) and the new (trend-based) similarity (6), where the weights have been determined empirically:

$$\mathrm{VS}_i(x_i, y_i) = 0.3\ \mathrm{VS}'_i(x_i, y_i) + 0.7\ \mathrm{VS}''_i(x_i, y_i) \tag{7}$$

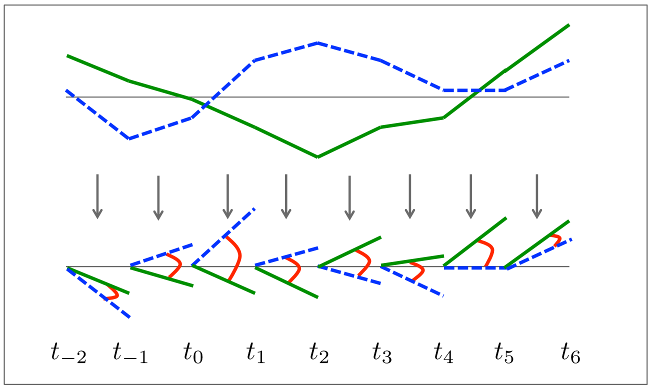

Fig. 3: Representation of time series as a bundle of vectors. The similarity for each pair of vectors depends on the angle between them (0 for an angle of π, 1 for an angle of 0). These similarities are averaged to obtain the overall similarity.

4.2 Case similarity

According to (1), the similarity between two cases is defined as a weighted average of the variable similarities. In previous work, the flexibility of weighting has actually not been exploited, i.e., all weights were simply set to the same value $v_i = 1/M$. However, since different variables are obviously of different importance, a generalization of this approach is desirable.

The determination of optimal variable weights v_i is closely connected to the problem of learning similarity measures, which has been studied intensively in CBR [21, 22, 11]. More specifically, the problem is to optimally combine given local (variable) similarities into a global (case) similarity [6]. To solve this problem, we take advantage of the fact that, according to (1), the combination is a *linear* one, i.e., global similarity is a linear (convex) combination of local similarities.

Concretely, we formalize the problem of learning weights v_i for variables V_i as a problem of *linear regression*: For every pair of cases X and Z from our case base, we can compute the (local) variable similarities

$$(s_1, \ldots, s_M) = \Big(\mathrm{VS}_1(x_1, z_1), \ldots, \mathrm{VS}_M(x_M, y_M)\Big) \in [0, 1]^M \ .$$

Moreover, we can compute a similarity s_{out} on the consumptions measured for X and Z, again using the transformation (2) of their Euclidean distance, with proper choices of d_{min} and d_{max}.[6] Ideally, the (global) case similarity is close to this value, i.e., $s_{out} \approx \mathrm{CS}(X, Z)$. Therefore, the weights v_j in (1) should be such that

$$\sum_{j=1}^{M} v_j \cdot s_j \approx s_{out} \quad . \tag{8}$$

As already said, an (approximate) equation (8) can be derived for each pair of cases from the case base, and each such equation can be seen as a training example for a (multivariate) linear regression problem, with the values of the input variables given by $(s_1, \ldots, s_M)$ and the value of the output variable by s_{out}. Thus, optimal weights can simply be found by solving this regression problem; more specifically, since the weights, which correspond to the regression coefficients, must be non-negative and sum up to 1, a constrained regression problem needs to be solved.

4.3 Adaptation

According to (4), similar cases retrieved from the case base are used in the prediction step without any adaptation. As a potential improvement, we propose a method for adaptation that is inspired by the idea of amplitude transformation [7]. More specifically, assuming that the future relation of energy consumption for two cases will approximately equal the relation in the past, the energy consumption of a source case retrieved from the case base is shifted by a proportional factor prior to using it for prediction.

Recall that the values q_{-2}, q_{-1}, q_0 for electricity consumption are assumed to be known for the query case (while consumption needs to be predicted for the six hours ahead), and let $p_{k,-2}, p_{k,-1}, p_{k,0}$ denote the consumption of the k^{th} neighbor in the past three hours. We then replace each of the future values $p_{k,j}$ $(j = 1, \ldots, 6)$ of that case by

$$p_{k,j} \cdot \left(\frac{q_{-2} + q_{-1} + q_0}{p_{k,-2} + p_{k,-1} + p_{k,0}} \right)$$

before using it for prediction in (4); see Figure 4 for an illustration.

4.4 Other modifications

Instead of retrieving all past cases with a similarity $\mathrm{CS}(X, Y)$ exceeding a fixed similarity threshold (of 0.8), we fix the number of neighbors to be used for prediction to $K = 50$ and retrieve the K most similar ones (if there are less than K cases with a similarity > 0, these cases are all retrieved).

[6] We used $d_{min} = 15$ and $d_{max} = 35$.

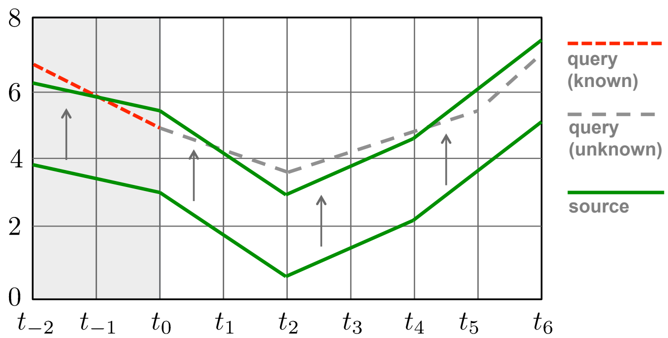

Fig. 4: Adaptation: The source sequence (solid line on the bottom) is shifted upward, so that the mean in the past (first three time points, gray region) coincides with the mean on the query (dashed line); the future values of the query need to be predicted and are therefore shown in gray.

Table 3: Variable thresholds and weights

Variable	d_{min}	d_{max}	weight
Forecast outside air temperature	2	6	0.1961
Forecast outside air relative humidity	10	25	0.1540
Air handling unit 2 supply hot air temperature	2	6	0.0001
Air handling unit 3 supply hot air temperature	2	6	0.0001
West wing air handling unit supply cold air temperature	2	6	0.009
Air handling unit 4 supply cold air temperature	2	6	0.1065
Chiller outlet water temperature	2	15	0.1075
Chiller outlet water flow rate	5	30	0.0001
Boiler outlet water temperature	2	15	0.3064
Boiler outlet water flow rate	5	30	0.1284

5 Experiments

5.1 Data

Data was collected from an institutional building facility located in Calgary (Alberta, Canada) for working days between 1^{st} of January 2013 and 9^{th} of May 2014 (with some missing data from 29^{th} of March to 1^{st} of May). The building has a total floor space of 16,800 m^2 and houses mainly office and storage spaces. The HVAC equipment consists of 5 air handling units served by a one chiller and 3 natural gas boilers. The data consists of hourly averages of measurements related to the operation of the chiller, boilers and air handling units, the building electricity consumption, and weather information—current and forecast values of outside air temperature and relative humidity (see list of variables in Table 1). Building operating modes corresponding to office working and non-working hours were identified. The building consumes approximately 80% more electricity during working hours—7 AM to 5 PM—than during non-working hours; only the model developed using working-hours measurements is presented in this paper.

5.2 Methods

Our CBR approach was implemented as described in the previous section. Under certain circumstances, it may happen that the case base does not contain a single case that is similar (i.e., has a similarity > 0) to the query case. In such a situation, our method yields the current consumption (i.e., the consumption q_0 at time t_0) in the query case as a default prediction for $\hat{q}_j$ for the next time points ($j = 1, \ldots, 6$); this predictor will also be used as one of our baselines (see below). For learning the weights of variables in the case similarity measure (cf. Section 4.2), we constructed a set of training data by randomly sampling 10,000 pairs of cases from the case base. The weights obtained by linear regression, which are shown in Table 3, are plausible and indeed give the highest importance to those variables that are intuitively deemed most relevant.

We compare our CBR approach with a number of other methods that are used as baselines to compete with. The first three baselines are extremely simple, and they all forecast a constant value for the six hours prediction horizon. They predict, respectively,

- the average consumption of all past cases stored in the case base;
- the average consumption $(q_{-2} + q_{-1} + q_0)/3$ of the past three hours in the query case;
- the current consumption q_0 in the query case.

Moreover, following [19], we also included an artificial neural network (ANN), namely a multilayer perceptron with one hidden layer consisting of 10 neurons, trained using the back propagation algorithm with Levenberg-Marquardt optimization. As input, the network takes the measurement values of the current and past two hours of a case, including the energy consumption (hence 33 values in total), and as output, it produces predictions of the energy consumption for the next six hours.

5.3 First experiment

In the first experiment, the data is separated into two parts: a set of past cases with measurements from the first m months of 2013 (where $m \in \{4, 6, \ldots, 12\}$) that corresponds to our case base and serves as training data for the ANN, and the remaining set of future cases till September 2014 that serves as test data. Performance is reported in Table 4 in terms of the CV-RMSE (5) on the test data. As can be seen, our CBR approach compares quite favorably and is much better than the baselines. The performance of the ANN is even slightly better if enough training data is available, but CBR seems to have advantages if training data is sparse.

Table 4: Results of the first experimental study in terms of CV-RMSE (%), with the best performance highlighted in bold font.

training data	baseline 1	baseline 2	baseline 3	ANN	CBR
01/2013 – 04/2013	9.97	9.55	8.80	8.99	**7.94**
01/2013 – 06/2013	10.15	9.50	8.65	8.15	**7.39**
01/2013 – 08/2013	10.23	9.41	8.58	7.63	**7.45**
01/2013 – 10/2013	10.12	9.35	8.66	**7.31**	7.69
01/2013 – 12/2013	10.35	9.54	8.73	**6.17**	6.55

5.4 Second experiment

In the second experiment, we applied our CBR approach in an online setting, in which prediction and learning (case memorization) are interleaved: Cases are considered in a sequence one by one, and at each time step t,

- a prediction of the consumption for the t^{th} case is obtained based on the previous $t-1$ cases already stored in the case base,
- the true consumption is revealed, and the cumulative error (CV-RMSE on the first t cases) is updated,
- the new case is added to the case base.

As can be seen in Figure 5, the performance is relatively poor in the beginning, when only few cases are available, but quickly improves and then reaches a level similar to the error (around 6.3%) in the previous experiment. This is a significant improvement compared to the previous CBR approach, for which the error is twice as high [19, 20].

6 Conclusion

This paper presents the application of a CBR model for predicting the hourly electricity consumption of an institutional building located in Calgary, Canada. The model uses measurements related to the building operation, as well as measured and forecast weather information to predict the building electricity consumption for the next 6 hours. It is based on a previous CBR approach applied to the same problem, however, modifications and extensions related to variable and case similarities, case selection and adaptation resulted in significant predictive accuracy improvements: The model has a test error approximately twice as low compared to the previous approach. This is important, as predictive accuracy is critical in enabling operators to take the appropriate operation and control decisions that ultimately result in reduced building energy consumption.

There are several directions to be pursued in future work. First, there is probably still some scope to further improve predictive accuracy. Perhaps more interestingly, however, we also plan to apply the approach of *credible case-based inference* [13], which allows for predicting confidence intervals instead of only

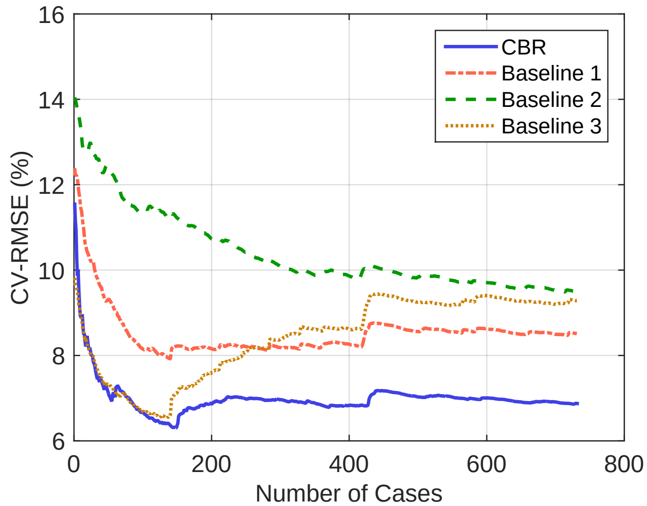

Fig. 5: Performance of the CBR model in the online setting.

point values. Thus, in our case, predictions will be intervals of the form $[q_t^{low}, q_t^{up}]$, coming with the guarantee that the true consumption will lie in that range with high probability. Predictions of that kind, reflecting uncertainty in a proper way, can usefully support safety-critical decisions, for example regarding peak loads.

Second, going beyond a single building, we plan to extend our approach toward knowledge (case) transfer between different building. As already mentioned, CBR appears to be especially suitable for realizing this kind of transfer learning, which, as a critical step, requires a reasonable approach to *inter-building* case adaptation in addition to the simpler *intra-building* case adaptation as presented in this paper. While hitherto results on single buildings, including those presented here, are certainly promising, we expect CBR to develop its true potential in that scenario.

Acknowledgments The first author gratefully acknowledges financial support by the EUROWEB+ program. The second and fourth author have been supported by the German Research Foundation (DFG), and the third author by the Program of Energy Research and Development (PERD) operated by Natural Resources Canada (NRCan).

References

1. International Energy Agency. Transition to sustainable buildings. Paris, 2013.
2. ASHRAE. Guideline 14-2002, Measurement of Energy and Demand Savings, 2002.
3. Continental Automated Buildings Association. North American Intelligent Buildings Roadmap. Ottawa, 2011.

4. A. Azadeh, S.F. Ghaderi, and S. Sohrabkhani. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management*, 49(8):2272 – 2278, 2008.
5. CABA. North American Intelligent Buildings Roadmap, 2011.
6. W. Cheng and E. Hüllermeier. Learning similarity functions from qualitative feedback. In K.D. Althoff, R. Bergmann, M. Minor, and A. Hanft, editors, *Proceedings ECCBR–2008, 9th European Conference on Case-Based Reasoning*, pages 120–134, Trier, Germany, 2008.
7. Kam Wing Chu, Kelvin and Man Hon Wong. Fast time-series searching with scaling and shifting. In *Proc. PODS'99, 18th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '99, pages 237–248, New York, NY, USA, 1999. ACM.
8. Drury B. Crawley, Jon W. Hand, Michal Kummert, and Brent T. Griffith. Contrasting the capabilities of building energy performance simulation programs. *Building and Environment*, 43(4):661–673, 2008.
9. D. Dohare and V. S. Devi. Combination of similarity measures for time series classification using genetic algorithms. In *IEEE Congress of Evolutionary Computation (CEC)*, pages 401–408, June 2011.
10. Yonggui Dong, Zhaoyan Sun, and Huibo Jia. A cosine similarity-based negative selection algorithm for time series novelty detection. *Mechanical Systems and Signal Processing*, 20(6):1461–1472, 2006.
11. Thomas Gabel and Eicke Godehardt. Top-down induction of similarity measures using similarity clouds. In *Proc. ICCBR 2015, 23rd International Conference on Case-Based Reasoning*, pages 149–164, Frankfurt am Main, Germany, 2015.
12. Pedro A. Gonzlez and Jess M. Zamarreo. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, 37(6):595–601, 2005.
13. E. Hüllermeier. Credible case-based inference using similarity profiles. *IEEE Transactions on Knowledge and Data Engineering*, 19(5):847–858, 2007.
14. Soteris A. Kalogirou and Milorad Bojic. Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy*, 25(5):479–491, 2000.
15. Danielle Monfet, E Arkhipova, and D Choiniere. Evaluation of a case-based reasoning energy prediction tool for commercial buildings. 2013.
16. Danielle Monfet, Maria Corsi, Daniel Choinire, and Elena Arkhipova. Development of an energy prediction tool for commercial buildings using case-based reasoning. *Energy and Buildings*, 81:152–160, 2014.
17. Dongxiao Niu, Yongli Wang, and Desheng Dash Wu. Power load forecasting using support vector machine and ant colony optimization. *Expert Syst. Appl.*, 37(3):2531–2539, March 2010.
18. Ping-Feng Pai and Wei-Chiang Hong. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*, 74(3):417–425, 2005.
19. Radu Platon, Vahid Raissi Dehkordi, and Jacques Martel. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. *Energy and Buildings*, 92:10–18, 2015.
20. Radu Platon, Jacques Martel, and Kaiser Zoghlami. CBR model for predicting a building's electricity use: On-line implementation in the absence of historical data. In *Proc. ICCBR 2015, 23rd International Conference on Case-Based Reasoning*, pages 306–319, Frankfurt am Main, Germany, 2015.

21. A. Stahl. Learning similarity measures: A formal view based on a generalized CBR model. In *Proc. ICCBR–05, International Conference on Machine Learning*, pages 507–521, Chicago, USA, 2005. Springer.
22. A. Stahl and T. Gabel. Optimizing similarity assessment in case-based reasoning. In *Proc. AAAI–06, 21st National Conference on Artificial Intelligence*, 2006.
23. Hai-xiang Zhao and Frdric Magouls. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6):3586–3592, 2012.