Using Case Based Reasoning to Predict Energy Intake for a Photovoltaic Solar Panel in a Non-stationary Environment

Anders E. Braten, Abdulmajid Murad

Department of Information Security and Communication Technology
Norwegian University of Science and Technology, NTNU
Trondheim, Norway

Abstract

Many IoT devices harvest energy from the environment to preserve their energy buffer and increase operating time. When operating in an environment where conditions change drastically over time, it is useful to have the ability to predict the potential future energy production. Case Based Reasoning (CBR) is a possible method for identifying how much energy we can expect to harvest for a given day, assuming we have stored knowledge of the energy output for a given device, and the conditions that was present at the time. In this paper, we describe a case study where the purpose was to investigate the appropriateness of using CBR for this purpose. Our results shows that it is possible to use CBR to predict the energy intake of a monitored photovoltaic solar panel, by using data collected from a public weather forecast and the calculated position of the sun.

Keywords: Case-based reasoning, solar powered devices, predictive model selection

1. Introduction

Having the ability to preserve their energy buffer is crucial for constrained, wireless nodes. One way of doing this is to exploit energy harvesting mechanisms to make the energy budget of the device neutral over time, i.e., make sure the energy intake is the same or higher than the energy consumption. To be sure that the energy budget is in balance, it is often necessary to adjust energy consumption to the identified future energy harvesting production. This require some sort of prediction algorithm for energy intake.

Energy harvest from a photovoltaic (PV) solar panel is largely dependent on two factors: Sun angle and cloudiness [1]. The position of the sun is easily calculated, while cloudiness can be obtained from public weather forecast services. There are traditionally three different approaches that are used to predict future energy intake for a device. The first is by calculating the energy output using a physical clear-sky model, and then subtract a given factor based on the anticipated level of cloudiness. The second method is based on a smart persistence model, which assumes that the output 24h ahead will be roughly the same as the current output, adjusted by the difference in sun elevation. The third way of predicting future energy intake is to use a statistical model, for instance by training a machine-learning algorithm using the sun position and cloudiness as input features [2].

However, we want to investigate if case based reasoning (CBR) can be used for this purpose as well. The intuition is that for a given sun position, a PV-panel will produce roughly the same amount of energy under similar weather conditions. That means that if we have knowledge of the sun position and weather conditions for a certain day, in addition to how much energy a specific PV-panel produced under those conditions, we can organise that knowledge as cases and use CBR to assess similarity between them.

In this paper we conduct a case study where we use the calculated position of the sun, weather forecasts collected from a public weather service and monitored data from a PV-panel mounted locally to construct a case base, where each day represents a single case. We then retrieve the seven most similar days, and predicts the energy intake of the PV-panel using the weighted average of the retrieved case set. Our initial results shows that CBR is a plausible and promising method to use for this kind of predictions.

This paper is organized as follows: In Section 2 we present related work. In section 3 we describe the data collection process, the case base construction, and the definitions of feature weights and similarity functions. In Section 4 we show our results, which are then discussed in Section 5. Finally, we give a conclusion in section 6.

2. Related work

Research on case based reasoning related to energy harvest prediction is limited. However, some papers discuss the usage of CBR for predicting energy consumption in buildings.

In [3], Monfet et al. discuss an approach to predict the energy demand of commercial buildings using case-based reasoning. Their approach is based on hourly predictions of the energy consumption for the next 3 hour interval, with weather forecast input features. The CBR method was evaluated using real data collected from an office building in Quebec, Canada. This case study shows that it is possible to use CBR for energy demand prediction.

In a study by Platon et al., CBR is used in combination with artificial neural networks (ANN) to develop simplified, yet accurate models that predict the electricity consumption of an institutional building. The models use a relatively small number of variables related to the building operation, in addition to weather forecast information, as input features. As in the previous case study, the actual predictions were carried out on an hourly basis, this time with a horizon of one to six hours. By using principal component analysis (PCA) the authors were able to reduce the number of inputs without decreasing the model accuracy. When comparing the predictive performances of the models, the results showed no accuracy loss when using only PCA-selected inputs [4].

In [5], Gonzlez-Briones et al. presents a Multi-Agent architecture (MAS) for a wireless sensor network (WSN) that aims at achieving energy savings in an office building. Here, CBR is used to predict the working hours of the employees, while the agents of the MAS learn social behavior by the use of an artificial neural network (ANN). In addition, the system has knowledge of the weather forecast for the coming days and periods where there are less activity, like holidays and non-working hours. This way, the system is able to set radiators and air conditioning optimally in regard to energy consumption.

3. Constructing the case base

3.1. Defining a case concept and main features

Using CBR for prediction require that we construct a case base that contain the knowledge that is necessary to find similarities [6]. We also need methods to assess similarity and retrieve cases. To this end we use the mycbr tool [7].

Since our aim is to identify the total energy produced by a specific pv-panel for a day under certain conditions, we choose to use a single day as concept for the cases. Each case, or day, has a set of features, represented by attribute-value pairs. These describe the position of the sun, the weather conditions and the energy production for that day.

The energy that can be produced by a fixed photovoltaic (PV) panel depends for the most part on how much sunlight that hits the panel, and the angle between the beam of the light and the plane of the panel. The angle between the sun and the panel is based on the sun position as seen from a given location, at a given point in time. The sun position follows a pattern that goes from a minimum at the winter solstice to a maximum at summer solstice, as shown in Fig.1. This implies that each day, the sun position will have a maximum elevation above the horizontal plane that is slightly different from one day to the next. We choose to use this number to describe the sun position for a given case.

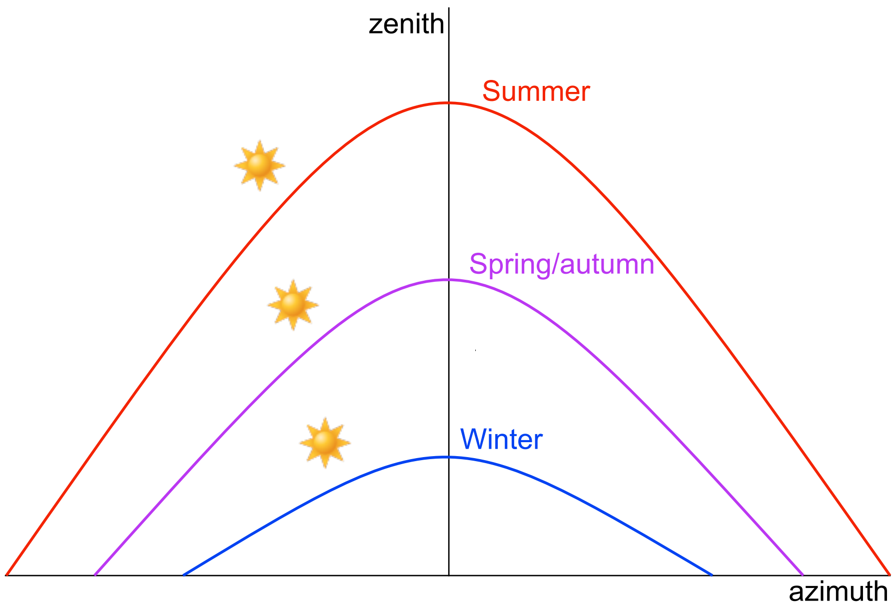

Figure 1: The apparent position of the sun during a day, as seen from a fixed location in the upper, northern hemisphere

Assuming that there are no shadows, the amount of light that actually reach the PV panel depends for the most part on how much of the light is blocked or scattered by clouds. The effect of clouds can be seen in Fig. 2. Here the actual energy output from a small pv-panel is charted over a course of three consecutive days. We can see that in this period, the max energy output on a day with heavy clouds is roughly 1/8th of what is produced

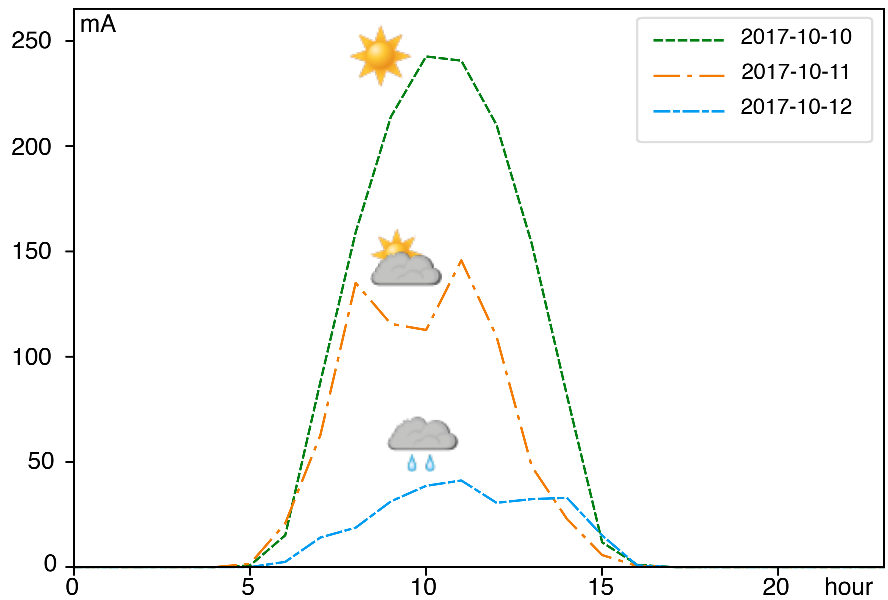

Figure 2: Actual energy production from a small pv-panel located in Trondheim, on three consecutive days in October 2017

on the day with a clear sky. As a measure of how much clouds there are in the sky, we can use cloudiness. Cloudiness is given as a value between 0 and 100, where 0 means that sky is completely clear, and 100 means the sky is completely obscured by clouds. A challenge is that this measure does not say anything about density, which imply that heavy clouds, that blocks more of the sunlight, are not given any more weight than light clouds, which block and scatter the light to a lesser degree. To compensate for that, we define cloudiness for three levels of altitude, i.e., high, medium and low. However, since the amount of the clouds within each layer vary from one day to the next, while the effect as seen from the ground is based on the distribution in each layer, we want to include a measure of total cloudiness as well. That will ensure that we can find similarity between two days with same amount of total cloudiness, even though the clouds are found in different layers.

In winter, very dense clouds may affect the number of samples that is collected, since too little light reach the PV-panel to activate energy production. We therefore add the number of samples as a feature to reflect this effect. Finally, since the temperature has some effect on the efficiency of the energy production, we also add that feature to the case.

The features described above are enough to assess the similarity between days. However, since we want to use CBR to predict total energy produced during one day, we need to add total energy production as a feature too. This feature will also be used to evaluate the accuracy of the prediction that is made. In addition, we add maximum energy produced, and mean energy produced. This makes it possible to analyse aspects of the energy production for a day. For instance, the ratio between maximum and mean energy produced might say something about the volatility of the weather for the day. On a normal day with stable weather conditions we expect the mean to be around 50 % of maximum. If the mean is low compared to max, this is an indication that there were few clouds in the middle of the day, but cloudy in the morning or evening. On the other hand, if the mean is close to max, the sky most probably was clear in the morning or afternoon, but obscured by clouds in the middle of the day.

Table 1: Description of features included in the case.

Feature	**Description**	**Weight**	**Polynom.**
Input features			
Max sun elev.	Max. angle between the horizon and sun position	8.0	12
$\text{Cloudiness}_{\text{low}}$	Mean cloudiness at altitude < 2000 m	3.0	5
$\text{Cloudiness}_{\text{medium}}$	Mean cloudiness at altitude of 2000-6000 m	2.0	5
$\text{Cloudiness}_{\text{high}}$	Mean cloudiness at altitude > 6000 m	1.0	5
$\text{Cloudiness}_{\text{total}}$	Mean cloudiness at all altitudes	2.0	5
Mean temp.	Mean temperature for the day	2.0	10
Sample count	Number of samples collected for one day	1.0	8
Output features			
Total energy	Total energy produced during one day	n/a	n/a
Max energy	Maximum energy registered during one day	n/a	n/a
Mean energy	Mean energy produced during one day	n/a	n/a

3.2. Adding instances to the case base

To populate the case base with instances, we collect data from various sources. First, the max sun elevation for each day is calculated using the astral library in Python. The four different values of cloudiness and the temperature is taken from a data set that we have collected in real-time from the public weather forecast service api.met.no. Finally, the three energy production values are calculated from another data set, which is collected in real-time from a local PV-panel that is mounted in plane with the horizon. These three data sets are then merged, using the timestamp as common denominator, and combined into one single dataset. The sample rate of this combined data set is one observation every 10 minutes.

Before we add the data to a case base, we do some data cleaning and preparations. First, we remove all observations where the energy production is 0. We then calculate the number of samples taken in a day, and take the nominal mean of all the values. This results in one observation for each day.

As mentioned in Section 3.1, the maximum sun elevation in a day goes from a minimum at winter solstice to a maximum in summer solstice. From this it follows that between these two days the maximum sun elevation will be at approximately the same elevation two times of the year. To reduce the number of cases, we therefore select three periods in the dataset where the maximum sun elevation for a day is in the range from 15.5 to 38.5 degrees above the horizon. This corresponds to 20th of February to 20th of April 2018, 20th of August to 20th of October 2018, and 20th of February to 20th of April 2019. Finally, we remove all days where the total energy production is less than 1500 mA, since this is an indication that there were some anomaly in the observation, like snow covering the PV-panel. In total, this cleaning resulted in a case base populated with 174 instances.

3.3. Setting the global and local similarity measures

To measure similarity between two cases, we need to define how similar two cases are on global and local level.. The overall goal of similarity modeling is to provide a good approximation that is reusable across cases and is easy to compute [8]. Each case is represented as a feature vector, i.e., a set of attribute-value pairs. This makes it possible to define a local similarity measure for each feature. The global similarity measure is calculated as a weighted average of all the local similarities.

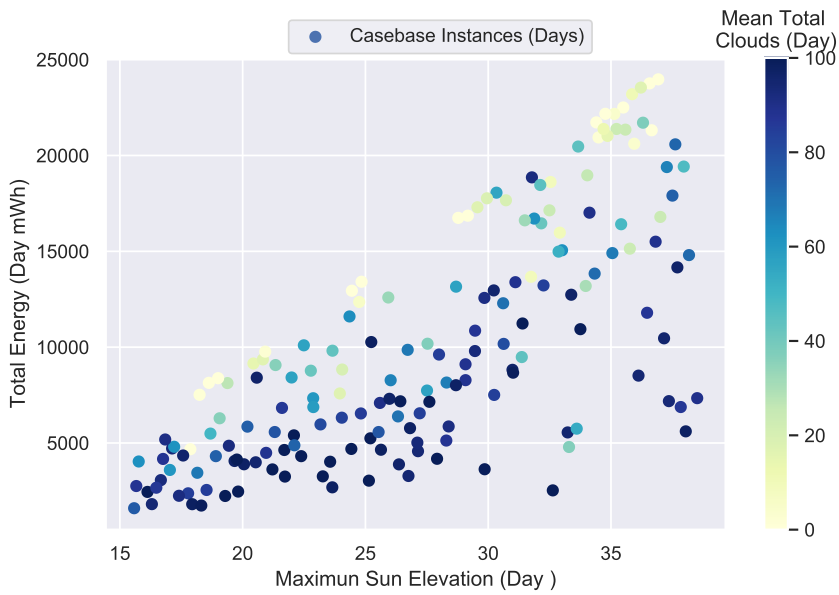

Figure 3: All cases in the case base with the attributes effect of sun elevation and clouds on the total energy production

The effect of sun elevation and cloudiness on total energy consumption can be seen in Fig. 3. From the diagram, we see that sun elevation is the dominant factor for energy production in a PV-panel, and is thus given most weight in the global similarity function. We give cloudiness$_{\text{low}}$ the second highest weight, since clouds closer to the ground, usually blocks more light than clouds found at higher altitudes. Conversely, cloudiness$_{\text{high}}$ is given the least weight. The weights of all features are shown in Table 1.

The local similarity functions that is used for the actual comparison between the instances in the case base is based on polynomial functions, i.e., the similarity increases with the closeness of the query. For each feature we compute a similarity between 0 and 1. The different polynomials used to calculate local similarity is given in Table 1.

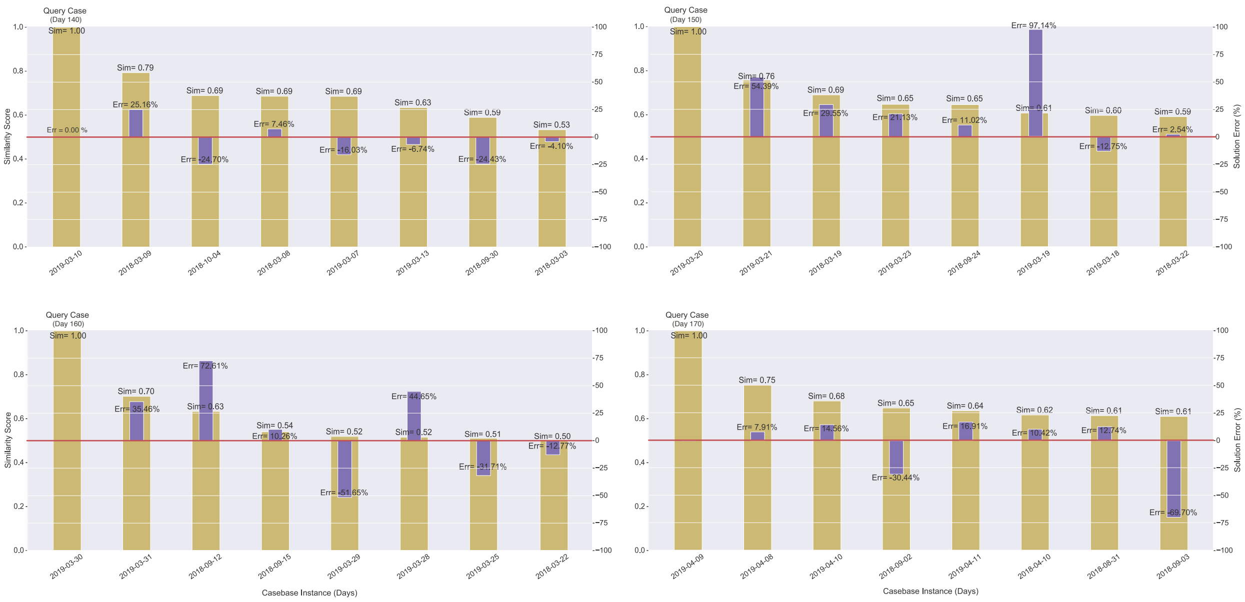

Figure 4: Most similar retrieved cases for four query cases (Day 140, 150, 160 and 170) showing similarity scores and solution error for every retrieved case. (Larger versions of these diagrams are provided in Appendix A.)

4. Results

4.1. Retrieving cases and make predictions

This section presents the results of the retrieval process on the built case base. In this step the most similar cases are retrieved based on the similarity of attributes to a query case to predict the solution, which is the total energy produced in a day. We evaluate the effectiveness the retrieval process by comparing the solution of the retrieved cases to the true solution of a query case.

Fig. 4 shows four examples of the retrieval process. For each query case we retrieve the seven most similar cases, based on the local and global similarity measures we defined in Section 3.3. The retrieved cases are ordered by similarity score. The diagrams also show the solution error of the retrieved cases which is the normalised difference between the predicted total energy and true total energy produced in the query case. For comparison, Table 2 and Table 3 presents the numerical values of the attributes in the query and the retrieved cases for Day 140 and Day 160, respectively.

Fig. 5 shows the Self-Similarity Matrix where each case from the case base is queried and its similarity to other cases are plotted in a heat map.

Table 2: Retrieved cases for day 140, ordered by similarity

Case ID	**Date**	**Sun elev.**	**Cl.$_{high}$**	**Cl.$_{low}$**	**Cl.$_{med}$**	**Cl.$_{tot}$**	**Temp.**	**#samples**	**Sim.**	**Energy$_{tot}$**	**% error**
day140	20190310	22.38	50.01	94.77	94.82	99.92	10.48	69	1.00	4313.32	0.00
day17	20180309	22.09	62.35	93.28	74.99	99.83	9.43	68	0.79	5398.34	25.16
day105	20181004	21.72	51.81	98.76	85.42	99.89	17.69	67	0.69	3248.11	-24.70
day16	20180308	21.70	25.08	99.56	69.50	99.87	9.76	68	0.69	4635.25	7.46
day137	20190307	21.21	79.08	99.54	94.51	99.97	10.96	65	0.69	3621.94	-16.03
day143	20190313	23.56	75.13	100.0	99.54	100.0	12.31	72	0.63	4022.76	-6.74
day101	20180930	23.26	83.84	83.11	89.77	99.65	19.03	70	0.59	3259.52	-24.43
day11	20180303	19.76	70.09	99.85	99.17	100.0	9.86	63	0.53	4136.37	-4.10
							Energy$_{tot, weighted}$:	**4090.67**		**% error$_{weighted}$:**	**5.16%**

Table 3: Retrieved cases for day 160, ordered by similarity

Case ID	**Date**	**Sun elev.**	**Cl.$_{high}$**	**Cl.$_{low}$**	**Cl.$_{med}$**	**Cl.$_{tot}$**	**Temp.**	**#samples**	**Sim.**	**Energy$_{tot}$**	**% error**
day160	20190330	30.24	68.00	29.10	54.44	84.21	15.12	77	1.00	7508.36	0.00
day161	20190331	30.63	70.57	7.09	62.80	80.01	14.09	79	0.70	10170.69	35.46
day83	20180912	30.23	63.44	78.99	48.16	95.61	22.89	81	0.63	12960.38	72.61
day86	20180915	29.08	74.68	53.92	54.17	88.22	21.17	79	0.54	8278.40	0.26
day159	20190329	29.86	95.42	92.00	91.68	99.42	16.18	80	0.52	3630.55	51.65
day158	20190328	29.47	48.15	78.10	47.94	87.13	21.67	79	0.52	10860.52	44.65
day155	20190325	28.29	73.74	44.70	47.07	87.87	14.45	69	0.51	5127.61	-31.71
day30	20180322	27.21	67.70	24.22	0.00	84.05	15.32	78	0.50	6549.35	-12.77
							Energy$_{tot, weighted}$:	**8465.58**		**% error$_{weighted}$:**	**12.74%**

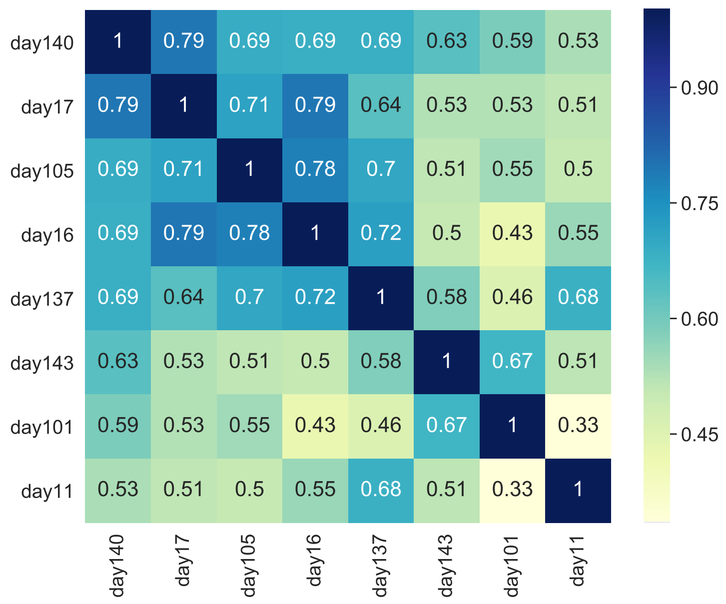

Figure 5: Self-Similarity Matrix for the query case (Day 140) and 7 most similar cases.

4.2. Prediction evaluation

To predict the total energy of a query case, we calculate the weighted average prediction of the retrieved cases as in equation 1. We use the similarity score to weight the retrieved predictions, where the most similar case contribute more to the solution.

$$Weighted\ Prediction_{query} = \frac{\sum_{ret=1}^{N_{ret}} Similarity_{ret} * Prediction_{ret}}{\sum_{ret=1}^{N_{ret}} Similarity_{ret}} \tag{1}$$

Then, to evaluate the quality of the prediction, we use the normalised absolute difference between the prediction and the true solution, as in equation 2.

$$Predition\ Error(\%) = \frac{\|True\ Solution_{query} - Weighted\ Prediction_{query}\| * 100\%}{True\ Solution_{query}} \tag{2}$$

To evaluate the overall prediction, we test every case as query case and calculat the weighted prediction of the retrieved cases, according to 1 for all cases. Then we calculate the prediction error according to 2. Figure 6 shows the retrieved prediction error for all cases in our case base when they are used as query cases, separately.

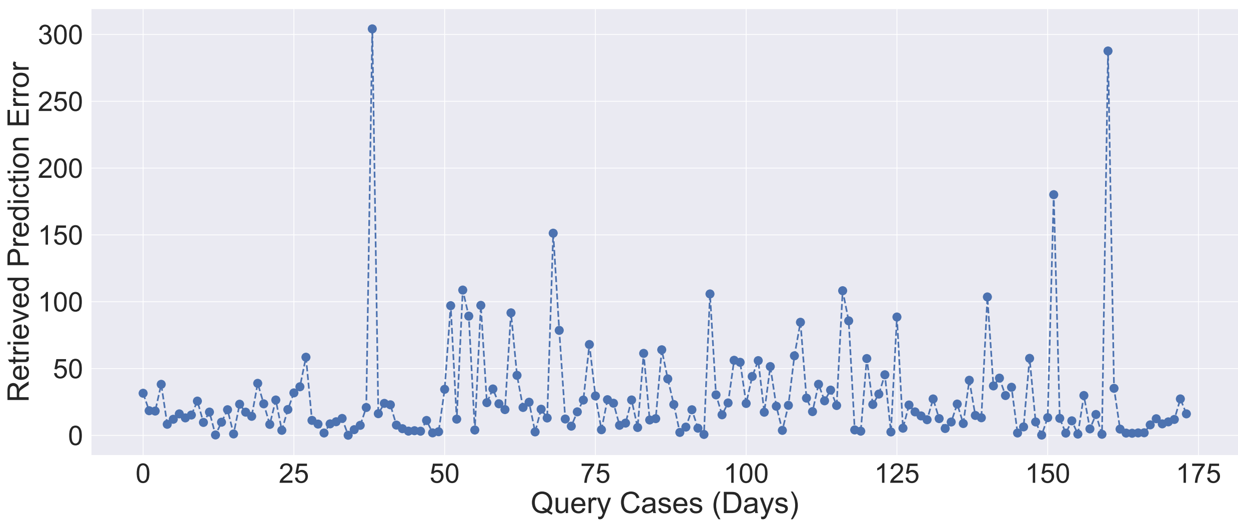

Figure 6: Retrieved prediction error for all cases in the case base when they are individually used as query cases

5. Discussion

The results show that CBR is a feasible method for predicting the energy produced by a PV-panel, given that we have a case base containing the necessary input- and output features, and look at a set of retrieved cases instead of just one. However, setting the local and global similarity measures require a high degree of domain knowledge, and also some knowledge of the weather patterns in the area where the PV-panel is mounted. Also, the similarity assessment is probably only valid for the specific type of PV-panel that are used to collect the necessary output feature, i.e., produced energy. This means that using CBR for energy prediction may not be easily generalised to other types of PV-panels or PV-panels located in an area with different weather patterns.

Fig. 4 and Fig. 6 shows that the solution error for each retrieved case has a high variance. If we now look at the Self-Similarity Matrix for day 140 illustrated in Fig. 5, we can see a trend where the similarity between case_{n} and $\text{case}_{\text{n}+1}$ decreases with the distance from the query case. An observation that is further supported by the data given in Table 2. These observations indicate that the case base may too small, i.e. for a given sun elevation we have few samples that shows the energy intake for different weather conditions. Thus, the prediction given by the case with the highest similarity has often low accuracy. However, this can possibly be mitigated by looking at the weighted average of the n most similar cases, as shown in our case study

The retrieval of similar cases are subject to seasonality. Since we haven't studied the effect of CBR-based predictions during the four months closest to the summer-and winter solstice, we can only speculate over the outcome for these two periods. The difference in sun elevation is smaller the closer to the solstice we are. This means, that the local similarity function for sun elevation will select a larger number of similar instances. Thus, given that the weather conditions for these instances have some variance, there will be more cases to assess for similarity. The intuition is therefore that this will increase the probability that we can find a day where the conditions where close to the query case, and thus are able to make more accurate predictions for the energy intake.

Finally, Fig. 6 shows the the prediction error given for each case. It should be noted that due to the implementation of the mycbr-tool that we used to retrieve similar cases, each case was tested for similarity against the whole case base. This means that our prediction algorithm was acting as an oracle, i.e., it had access to future information, which naturally would not be available in a real case. However, in a realistic deployment we can assume that the system would have access to a case base consisting of instances collected over a period of time prior to the time of deployment. Thus, under these conditions, the feasibility of using CBR as a method for energy intake prediction is still valid.

6. Conclusion

We have conducted an exploratory case study where we investigated the possibility of using CBR to predict the total energy intake produced by a photovoltaic solar panel during a day. Each day was represented by a set of attribute-value pairs, describing the sun position, weather conditions, and the energy produced under those conditions. By organising each day as a case in a case base, we were able to assess the similarity between days, using local and global similarity functions. Our results shows that by calculating the weighted mean of the seven most similar days to a query case, we were able to give good predictions of the energy production for the PV-panel that is monitored.

7. References

[1] A. E. Braten, F. A. Kraemer, Towards Cognitive IoT: Autonomous Prediction Model Selection for Solar-Powered Nodes, in: 2018 IEEE International Congress on Internet of Things (ICIOT), IEEE, pp. 118–125.

[2] F. A. Kraemer, D. Ammar, A. E. Braten, N. Tamkittikhun, D. Palma, Solar energy prediction for constrained IoT nodes based on public weather forecasts, in: Proceedings of the Seventh International Conference on the Internet of Things, ACM Press, New York, New York, USA, 2017, pp. 1–8.

[3] D. Monfet, M. Corsi, D. Choinire, E. Arkhipova, Development of an energy prediction tool for commercial buildings using case-based reasoning, Energy and Buildings 81 (2014) 152 – 160.

[4] R. Platon, V. R. Dehkordi, J. Martel, Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis, Energy and Buildings 92 (2015) 10 – 18.

[5] A. Gonzlez-Briones, J. Prieto, F. De La Prieta, E. Herrera-Viedma, J. M. Corchado, Energy optimization using a case-based reasoning strategy, Sensors 18 (3).

[6] A. Aamodt, E. Plaza, Case-based reasoning: Foundational issues, methodological variations, and system approaches, AI Communications 7 (1) (1994) 39–59.

[7] K. Bach, C. Sauer, K.-D. Althoff, T. Roth-Berghofer, Knowledge modeling with the open source tool mycbr, Vol. 1289, 2014.

[8] R. Lopez De Mantaras, D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T. Cox, K. Forbus, M. Keane, A. Aamodt, I. Watson, Retrieval, reuse, revision and retention in case-based reasoning, Knowl. Eng. Rev. 20 (3) (2005) 215–240.

Appendix A. Large versions of the diagrams showing similarity scores and solution error for the retrieved cases for Day 140, 150, 160 and 170

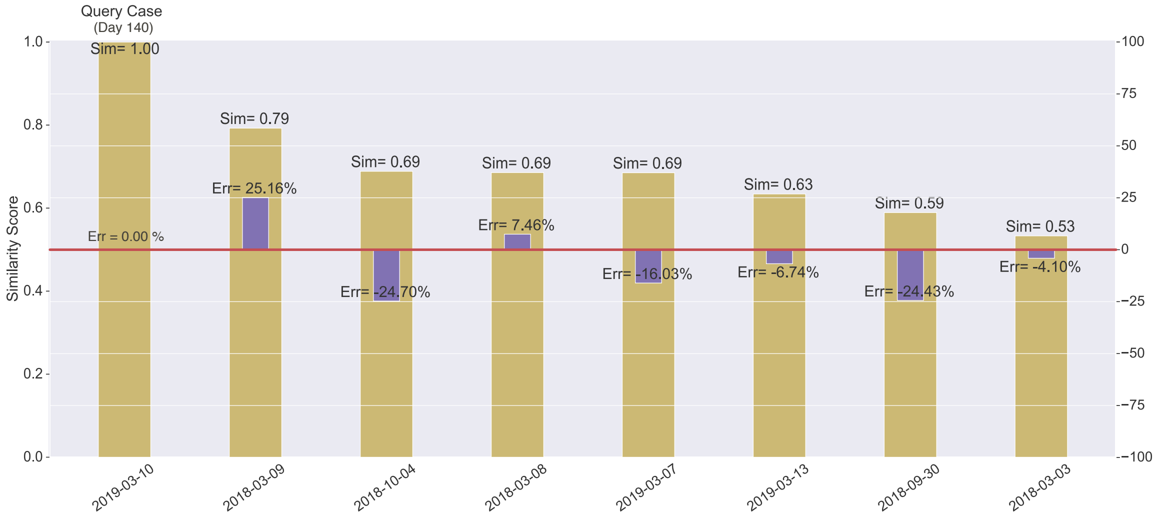

Figure A.7: Most similar retrieved cases for a sample query case (day140, 2019-03-10) showing similarity scores and solution error for every retrieved case.

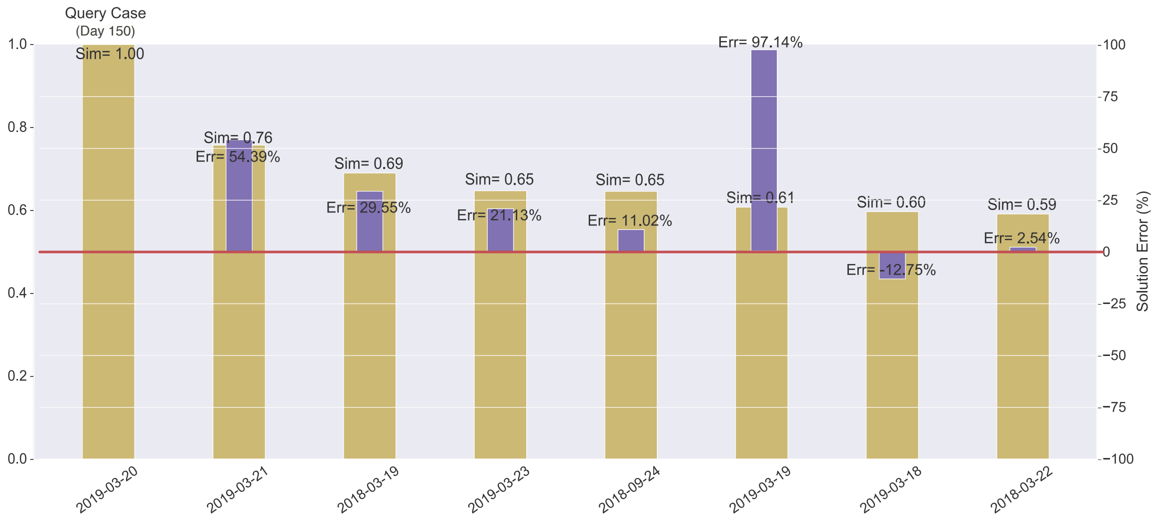

Figure A.8: Most similar retrieved cases for a sample query case (day150 ,2019-03-20) showing similarity scores and solution error for every retrieved case.

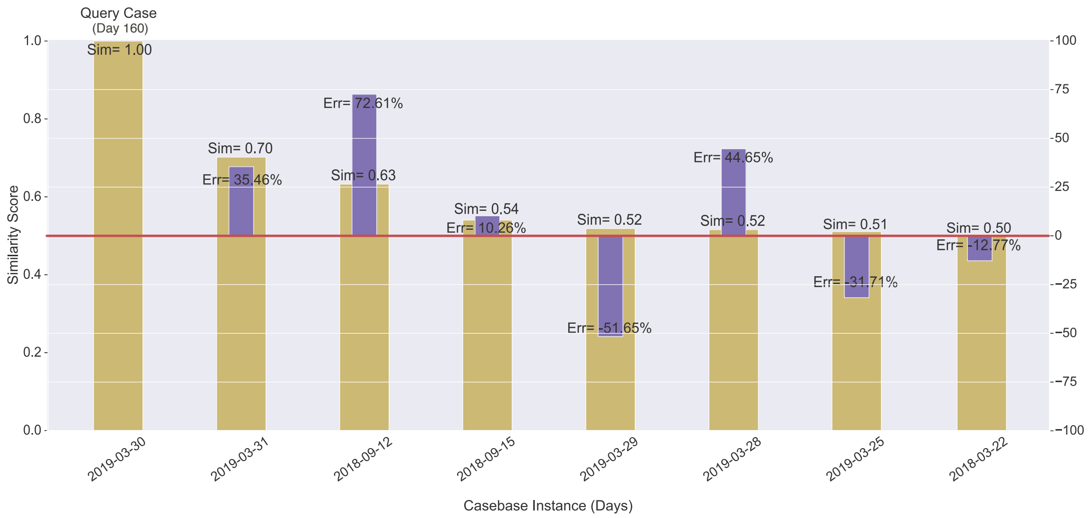

Figure A.9: Most similar retrieved cases for a sample query case (day160, 2019-03-30) showing similarity scores and solution error for every retrieved case.

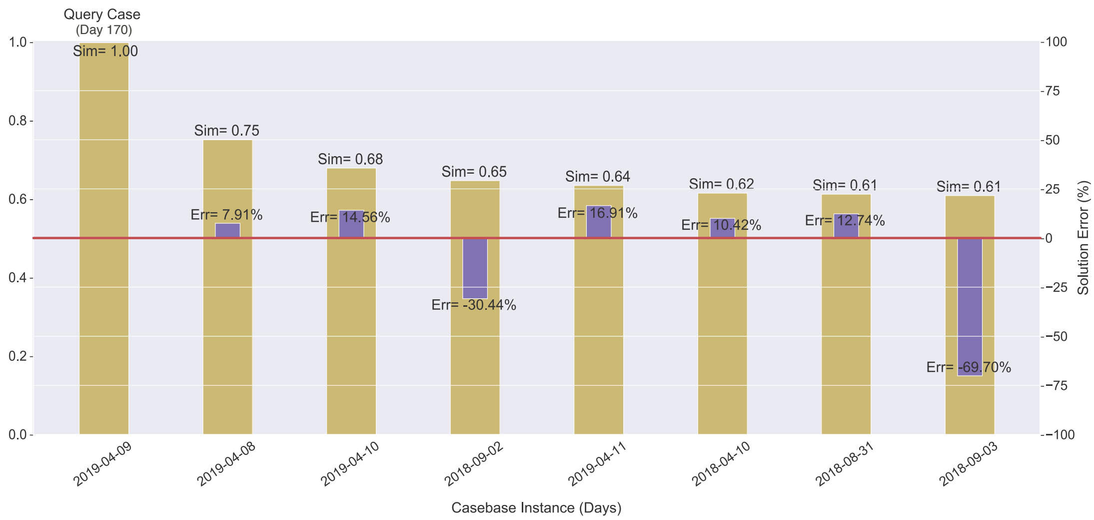

Figure A.10: Most similar retrieved cases for a sample query case (day170, 2019-04-09) showing similarity scores and solution error for every retrieved case.