Case Representation and Adaptation for Short-Term Load Forecasting at a Container Terminal

Norman Ihle

OFFIS, Institute for Information Technology, Oldenburg, Germany
norman.ihle@offis.de

Abstract. The electricity consumption of a terminal is mainly related to the number of container movements and the weather of each day. With the introduction of electric mobility for heavy duty container carriers at a seaport container terminal short-term load forecasting gains an important part in the procurement process. This paper describes a case-based approach to the forecasting of the electricity consumption time-series of the following day based on historical consumption load curves. It mainly focuses on the case representation which is based on a daily view on the so-called sailing list that is used to plan terminal operations and the adaptation processes that are applied to the time-series after case retrieval. The evaluation of the approach shows some promising first results.

1 Introduction

The introduction of electric mobility in industrial enterprises increases the importance of energy procurement even in sections where energy procurement was not in focus until today. Especially when some processes are changed from conventional to electric mobility in the division of logistics, economic potentials arise through the use of flexible electricity supply contracts. One example of such a logistic system is a seaport container terminal. Tasks like the transport of containers from the quay crane to the bulk storage area can be automated and electrified. This can be achieved by using battery-electric powered engines in the heavy-duty vehicles instead of diesel engines. In opposite to Diesel, that can easily be stored on terminal grounds, the fluctuating demand of the fuel 'electricity' has to be procured on a short-term basis. When using the electricity exchange, the day-ahead market is the last chance to match the procured electricity load curve with the latest demand forecast values. The forecast of a daily load curve, a time-series with 15-minute time stamps, is referred to as short-term load forecasting (STLF).

Over time many methods for short-term load forecasting have been developed and are now in use. Some of those methods are developed for specific scenarios, others follow a more general approach. Regarding the application of these methods to the load forecast of a container terminal there is no experience documented in the scientific literature. Even though the more general methods can

In Proceedings of the ICCBR 2016 Workshops. Atlanta, Georgia, United States of America

be applied to all kind of energy consumers, there is no systematic evaluation which method yields the best results when being applied to a single industrial site from the logistic domain. Since the consumption of electricity at a container terminal is subject to some special features that might not be available for other load forecasting scenarios, e.g. the consumption relies highly on the number of containers being handled, a solution based on Case-Based Reasoning (CBR) is proposed. The idea is that days with similar logistic operation requirements have similar load consumption patterns. This is due to the fact that only a few handling equipments have a great impact on the load profile, especially the quay and yard cranes and the reefer container.

In the following a case-based load curve forecasting approach is presented that is based on the list of expected ship arrivals and departures. In chapter 2 the case representation is introduced with a focus on how the data for exactly one day is transformed and represented. Chapter 3 describes shortly how time values are handled in regard to similarity. If a similar case is retrieved the corresponding consumption time-series can be adapted to fit it to the forecasting situation. Two approaches for the adaptation are introduced in chapter 4. Chapter 5 then presents the first results of the evaluation before chapter 6 discusses some related work. Chapter 7 closes the paper with the conclusion and an outlook on further work.

2 Case Representation

In seaport container terminals the number of logistic operations to be planned is very high. Containers are continuously delivered and picked up via train, truck or ship, the quay and yard cranes have to be assigned accordingly and the storage areas have to be administrated. The most important source for the operation planning is the list of ship arrivals and departures, the so called sailing list. The list is a table that includes the most relevant information for the expected ship arrivals and departures. The list is continuously updated with the newest available information that the terminal receives from the ship operators or freight companies. For example, a ship reports its departure from the previous terminal and the expected travel time to the next terminal. Due to weather influences on the traveling speed and other circumstances this information can vary throughout the route. One day before a ship arrives the expected arrival time is pretty reliable. Figure 1 shows some examples from the sailing list of the Container-Terminal Altenwerder (CTA) in Hamburg (Germany) from September 2013. The data from the sailing list can be used to build up the case base while the time-series data of the electricity consumption is stored in a dedicated database outside of the case structure due to performance and efficiency reasons. For simplification reasons one row of the sailing list will be referred to as *arrival* in the following.

JSNR	Ship Name	Ship Type	Expected Arrival	Expected Departure	Loading	Un-loading
308505	AKACIA	Feeder	04.09.2013 15:45	05.09.2013 04:00	373	244
306757	OOCL KAOHSIUNG	ATX	05.09.2013 00:10	05.09.2013 15:35	1399	16
308442	A LA MARINE	Feeder	05.09.2013 07:00	06.09.2013 06:00	534	556
308632	KAHN LAUK	Kahn	05.09.2013 15:55	05.09.2013 16:30	0	5
306926	EMOTION	Feeder	06.09.2013 07:45	07.09.2013 00:55	458	333
307896	APL VANDA	LOOP_7	06.09.2013 17:55	08.09.2013 14:00	2524	3031
308543	LEONIE P	Feeder	06.09.2013 15:45	06.09.2013 20:15	22	73

Fig. 1. Example of sailing list data

2.1 Using Sailing List Data for Case Modeling

In a first step data from the columns of the sailing list can be used for case modeling using an attribute-value representation. Besides the expected arrival and departure times of each ship, the sailing list includes information about how many containers have to be loaded and unloaded from each ship during the berthing time. These data together with the ship type can then be used to find similar days in the past. Similar in this case means that ships with a comparable number of containers to be handled have been arriving or departing at almost the same times of the day. The information is adopted for case modeling:

- JSNR: Attribute CS_{JSNR}
- Ship type: Attribute CS_{Type}
- Expected arrival time: Attribute $CS_{Arrival}$
- Expected departure time: Attribute $CS_{Departure}$
- Number of containers to unload: Attribute C_{Import}
- Number of containers to load: Attribute C_{Export}

The JSNR is a unique identifier of each ship's berthing time. The ship type distinguishes between a barge, a feeder ship and different seagoing vessels, often represented by the code of the regular tour they are operating on. The expected arrival and the expected departure times describe the expected berthing time, which is stored, expressed in minutes, in an additional attribute ($CS_{Berthing}$).

2.2 Splitting the Berthing Time to Represent One Day

Some entries in the sailing list represent a berthing that includes two or more dates according to the calendar, e.g. the ship *APL VANDA* with JSNR 307896 in Figure 1. Since the time span for the electricity consumption forecast is one day it is necessary to build up arrivals that represent the information for exactly one day from 00:00 to 23:59 o'clock. The available information has to be adapted to fit into a daily view of the sailing list information. To build up the daily view a berthing that exceeds the date limit can be split up to fit into multiple (virtual) arrivals, each representing the data for exactly one date. The information for JSNR 307896 can be split up to the following three entries:

– Arrival 06.09.2013 17:55 - Departure 06.09.2013 23:59
– Arrival 07.09.2013 00:00 - Departure 07.09.2013 23:59
– Arrival 08.09.2013 00:00 - Departure 08.09.2013 14:00

The number of containers to be handled per ship have to split up as well to fit the new virtual arrival information. For the split up, knowledge about the crane assignment at the seaside of the terminal can be used since the container handling is not constant during the berthing time. Using the simulation model presented in [2] and [3] it can be shown that the container handling numbers for large ships (more than 250 containers to be handled) decrease over the berthing time. After some preparation time the handling number starts at a high rate that is decreasing over time because cranes get allocated to other quay areas due to little space at the quay or higher priority of other arriving ships. To approximate this observation, a weighting function is introduced to calculate the number of containers C that is handled at each full hour h_i of the berthing time. Full hours are chosen here to approximate the distribution of the handling numbers over the time because more exact information is not available and a break-down to 15 minute values has no further advantage. Based on the average linear handling rate that is needed to load and unload all designated containers C_{total} of the ship during all full hours of the berthing time h_n it weights the first hours high and decreases linear over time:

$$C_{h_i} = (1,3 - \frac{0,6}{h_n - 1} * (h_i - 1)) * \frac{C_{total}}{h_n} \ with \ h_i = 1, ..., h_n \tag{1}$$

Based on this function the container numbers of a sailing list entry can be split into handling numbers for each virtual arrival and departure. If a ship arrives at the exact beginning of an hour, the first full hour reserved for handling preparations is not regarded for h_{total} and gets assigned a container handling number of 0. This is also applied if the ship arrives within a full hour.

2.3 Adding Additional Daily Information

Having adjusted the sailing list so that all available data can be assigned to exactly one day, daily cases can be built up. For each daily case additional information can be added. The values for the average temperature and wind speed of each day are of special interest because it can be shown that particularly the temperature value has an impact on the electricity consumption. Besides the weather information the weekday is added to the case since different repeating consumption patterns can be observed in the electricity consumption load curves of different days. While the differences between the weekdays are not as remarkable, differences between load curves of weekdays and days on the weekend differ even though the container terminal realizes a 24/7 operation. This might be explained by the lesser use of the office building and the Sundays driving prohibition for trucks in Germany. A special kind of weekday are holidays at which the terminal is actually shut down as well as the day before and the day after such a holiday. Figure 2 shows some typical load curves on different

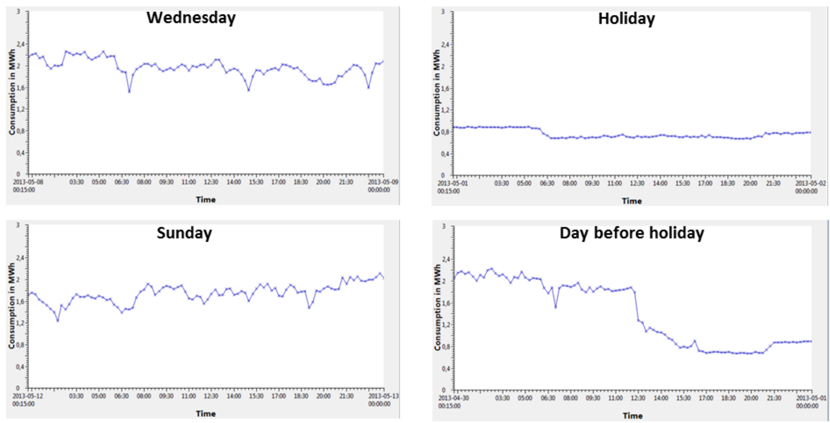

Fig. 2. Example of typical load curves

days. Two attributes with daily information can be derived from the sailing list information: the number of arrivals for a day and the overall number of containers that have to be handled. The second value is a good indicator for the average electricity consumption level for that day. The more containers have to be handled the higher the average level. Additionally an attribute saves the date value. It has the function to link to the corresponding time-series data of the day.

3 Similarity for Time-Based Attributes

When comparing two cases the arrival and the departure time of each ship plays an important role to assess the similarity. In the sailing list the information is available in a date-time format, the date being separated from the time by a blank. Since the single cases represent a daily view on the terminal operations, the date information can be disregarded: only the time information is interesting for similarity calculations. In order to be able to use simple existing similarity measures the time value is converted to an integer value. This value represents the number of minutes that have passed since the start of the day. The time value 00:00 therefore is represented by 0 and the end of the day at 23:59 by the value of 1439. If the absolute distance between the time values of two different arrival attributes is less than 60 both arrivals were not more than an hour apart at the point in time of the day and are therefore quite similar. The similarity decreases fast if the difference increases. To represent this, a sigmoid similarity function based on the distance is chosen as suggested by [1]:

$$sim_{CS_{Arrival}}(d_{CS_{Arrival}}) = \frac{1}{e^{\frac{d-120}{30}} + 1} \tag{2}$$

The same function can be used for the departure time as well and with slightly different parameters it can also be used for the attribute $CS_{Berthing}$.

4 Adaptation

The adaptation of the electricity consumption time-series related to similar cases is twofold. Both approaches are based on data stored in the cases. The first adaptation approach focuses on the single values of the time series and tries to adapt the values of each hour of one day based on the difference in the number of container to be handled in each hour. The second one is based on the differences in the weather data and is applied to the general level of all time series values for one day.

4.1 Hourly Container Number Comparison

It can be shown that the number of containers handled in an hour has a remarkable impact on the characteristics of the electricity consumption time series. The higher the number of containers handled the higher the electricity consumption. The idea behind the first adaptation approach is to find hourly differences in the number of containers handled on the day of the most similar case and the day of the query. For this purpose the aggregated container moves per hour i are calculated. For each arrival that is related to the day of $CASE_{sim}$ the container number is split into hourly values using the weighting function discussed in chapter 2.2. In this context not the exact daily values of the virtual arrivals are used, but the original entries are used in order to be able to use the weighting function properly and the hourly values are not stored. The same can be done for each arrival related to the day of the query ($QUERY$). Having done so, the difference d_i between the container handling numbers of $CASE_{sim}$ and $QUERY$ can be calculated for each hour. The first approach was to use this hourly difference directly to apply an adaptation factor to the time series values of that specific hour, but the evaluation showed that this approach had the shortcoming that it does not take into account the operations before and after the hour, but they also have an influence on the electricity consumption of that hour since preparations or post-processing also have to be executed. It might also happen that the difference in the container handling number is subject to strong fluctuations, for example when large seagoing vessels with a huge amount of containers arrive and depart, but the fluctuation in the electricity consumption is not as high at that point of time because of smoothing effects that cause some delay in the change of consumption pattern. For this reason not only difference d_i in the hour i is regarded when choosing an adaptation factor, but also difference in the hour before $i-1$ and the difference in the hour after $i+1$ are also considered. At the moment an adaptation factor is assigned to different times of day depending on the discussed differences using a rule set. Table 1 shows an extract of the rules as they are applied for the hour 8 (08:00 - 08:59 o'clock) of the time series. The $Factor$ is added to the time series values of the hour. It can be shown that the fluctuations are higher in the beginning of the day, so the adaptation factors are higher than the ones for morning and evening hours, while the adaptation factors for the hours in the middle of the day are rather low.

Hour i	d_i	d_{i-1}	d_{i+1}	$Factor$
8	$>$ 200	$>$ 150	$>$ 150	0.4
8	$>$ 150	$>$ 120	$>$ 120	0.3
8	$>$ 100	$>$ 70	$>$ 70	0.2
8	$>$ 50	$>$ 20	$>$ 20	0.1
8	$<$ -50	$<$ -20	$<$ -20	-0.1
8	$<$ -100	$<$ -70	$<$ -70	-0.2
8	$<$ -150	$<$ -120	$<$ -120	-0.3
8	$<$ -200	$<$ -150	$<$ -150	-0.4

Table 1. Adaptation factors for hour 8 of a day

4.2 Regarding Seasonal and Weather Influences

The second adaptation approach is based on the available weather information. The highest influence on the electricity consumption is based on the air temperature. It can be observed that the energy consumption of the container terminal is on average significantly higher in winter than it is in summer. This is mostly due to additional heating that is required when having low temperatures and more lighting is needed on winter days than in summer days. Since the available temperature values are average temperatures for one day, the differences in the temperatures can be applied to increase or decrease the level of the electricity consumption, but they are only supposed to be applied when the case represents a winter day and the query a summer day or the other way around and the temperature difference is greater than 5 degrees Celsius, since a difference of 5 degrees on summer days does not show a significant effect on the electricity consumption. Additionally boundaries for the adaptation were defined. This means that if a consumption value is already very high and the temperature based adaptation would raise this value beyond the upper boundary, the adaptation is not applied to avoid a forecast of values that are usually not reached under normal operating conditions of the terminal. The same applies to very low values and the lower boundary.

5 Evaluation

For the evaluation the sailing lists and the corresponding electricity consumption time-series of the years 2010 through 2013 of the Container-Terminal Altenwerder are used. Today, forecasts are generated by using the last week's values as forecast values (last week approach). While these values fit in some cases, they also generate some remarkable deviation from the forecast to the metered consumption in other cases. So besides improving the forecasting accuracy, it is also a goal to decrease the maximum deviations in the forecast compared to the real consumption. For the CBR approach the adapted electricity consumption values of the most similar case are used as forecast values. Besides this forecast a second one based on CBR is calculated by using the mean value of the adapted time-series values of the three most similar cases instead of only the one of the most similar case. This forecast will be referred to as CBR Top 3 approach. The first results of the evaluation show that the CBR approach can be successfully applied for forecasting the energy consumption of a container terminal. When forecasting the electricity consumption load curve for every day in June 2013, the last week approach yields an average Mean Absolute Percentage Error (MAPE) of 12.0 while the CBR approach yields to 10.0 in average. Having a yearly electricity consumption of more than 65000 MWh, a 2% better forecast avoids penalty fees that have to be paid when the procured electricity load curve does not fit the actual metered one can be avoided to a great extent. With the CBR Top 3 approach the results even yield to a mean of 8.6 which means a further improvement on the forecast. Figure 3 shows the MAPE values for every

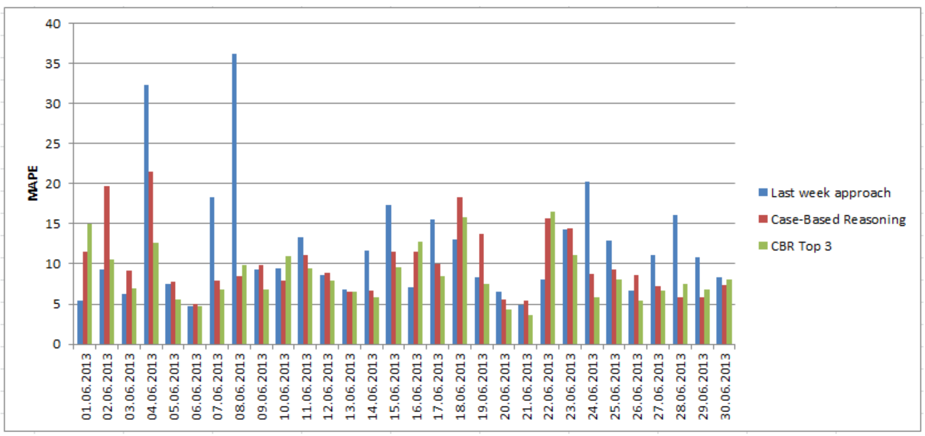

Fig. 3. MAPE values for June 2013

single day in June 2013 for the three approaches. It can be seen that on some days the MAPE values of the three approaches are close to each other while on some days the last week approach has very high values. On the 4th and the 8th of June the mean absolute percentage error of the last week approach is even higher than 30% while with the CBR Top 3 approach the highest values are slightly above 15%. This shows that the fluctuations can be decreased and the risks of over- or under-procurement of the electricity for a day is minimized. The Mean Squared Error (MSE) values of each day in June confirm these findings that can be observed when forecasting the following month as well.

Over the year 2013 the monthly average MSE of the CBR Top 3 approach is never higher than 0.07 while for the last week approach the maximum average value is double that value. The average MAPE for the year 2013 is 8,87% for the CBR Top 3 approach, while the currently used last-week approach has a MAPE of 12,32% in average.

6 Related Work

In the energy consumption forecasting domain only a few works have been published that present CBR approaches. [6] present the system FUTURA to forecast the medium-term load consumption in Peru. To adjust a calculated starting solution, the CBR module uses the average of the retrieved similar cases to smooth out fluctuations of the starting solution. The similarity is based on time attributes like month, weekday, and the time of day. Weather influences are disregarded. This adjusted solution is then adapted based with the expert knowledge regarding the long-term trends in environmental development. It is stated that the expert knowledge can change the load curve up to 20% depending on the influencing parameters. [7] presents an approach for the forecast of power peaks in a distribution grid based on CBR methods. As attributes the author uses weekdays, weekday-types and several weather indicators. The case base is clustered with methods of Self-Organizing Maps (SOM) to allow for a highly efficient retrieval. The k-nearest neighbor retrieval uses a weighted sum of all attributes of the request and a cluster. The most similar case is then retrieved out of the cluster with the highest similarity value. The power peak of this most similar day is then taken as power peak for the forecast day adapted by a factor that represents the yearly growth rate in the overall load of the grid. In a case study it is shown that the suggested solution outperforms simple last week value approaches and even artificial neural networks. [4] introduce a system for very short-term load forecasting of office buildings based on CBR-methods. The similarity is assessed with a focus on weather indicators like temperature, moisture, etc. The weights of the single attributes are dynamically adjusted regarding the current season and current weather conditions. Using a subset of the most similar solutions a forecast for the next three hours is built up. It includes a prognosis whether a new power peak will occur in this time horizon. The evaluation of the system shows that it can be applied successfully. The mean squared percentage error is stated to be around 12 to 14 percent. [5] extend this approach to a forecast horizon of 6 hours and present an on-line implementation of the system. They also specify the input attributes that are used. Besides the weather indicators and the current electricity consumption they also use power values of air conditioning and heaters currently in use and room temperatures currently metered within the office building. None of the related works deal with the daily short-term load forecasting of single industrial consumers based on operation plans.

7 Conclusion and Outlook

This paper presented an approach to use CBR methods to realize a system for short-term load forecasting for a single industrial customer, a container terminal, and focused on the case modeling and the adaptation of the time-series. The forecast of the electricity consumption for every quarter of an hour of the next day is based on the logistic operation plan that is set up for the next day. This plan includes data about expected ship arrival and departure times as well as information about the number of containers to be handled with each ship. This information can be used to find days in the past with a similar operation plan. The electricity consumption time-series of a similar day can than be adapted based on differences in the operations of each day and based on weather data. The approach has some advantages when compared to other established load forecasting methods: it is generally applicable to maritime container terminals, the computational cost is rather low and the results are comprehensible. The evaluations shows that the approach can be applied successfully.

Further improvements might be reached by integrating information about the container numbers that are delivered and picked up by truck or train to approximate the number of container handles and their course more exactly. At the moment this number is estimated because of lack of the data. For the adaptation process a yearly factor can be introduced that will represent the yearly average change in electricity consumption that can be observed over the previous five years. This is also important because in general the case-based forecasting approach tends to underestimate the actual values when applied at the use-case CTA, where the average yearly electricity consumption constantly increased during the last years. It is to be investigated if a CBR-STLF-approach for single electricity consumers based on operation plans can be generalized to further domains than container terminals. Logistic systems can be in focus here as well as more different systems like industrial production sites.

References

1. Bergmann, R.: Experience Management: Foundations, Development Methodology, and Internet-based Applications. Springer-Verlag, Berlin, Heidelberg (2002)
2. Grundmeier, N., Hahn, A., Ihle, N., Runge, S., Meyer-Barlag, C.: A simulation based approach to forecast a demand load curve for a container terminal using battery powered vehicles. In: Neural Networks (IJCNN), 2014 International Joint Conference on. pp. 1711–1718. IEEE (2014)
3. Grundmeier, N., Ihle, N., Hahn, A.: A discrete event-driven simulation model to analyse power consumption processes in container terminals. In: Simulation in Production and Logistics 2015. Fraunhofer IRB Verlag, Stuttgart (2015)
4. Monfet, D., Arkhipova, E., Choiniere, D.: Evaluation of a case-based reasoning energy prediction tool for commercial buildings. International Conference for Enhanced Building Operations (2013)
5. Platon, R., Dehkordi, V., Martel, J.: 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)
6. Vilcahuaman, R., Melendez, J.: Improvement of electric load forecasting reusing past instances and contextual based adaptation. In: Transmission and Distribution Conference and Exposition: Latin America, 2004 IEEE/PES. pp. 271–276 (Nov 2004)
7. Wang, Z.Y.: Developed case-based reasoning system for short-term load forecasting. In: 2006 IEEE Power Engineering Society General Meeting. pp. 6 pp.– (2006)