Poster: Visual Prediction of Time Series

Ming C. Hao, *Halldór Janetzko, Ratnesh K. Sharma, Umeshwar Dayal, *Daniel A. Keim, Malu Castellanos

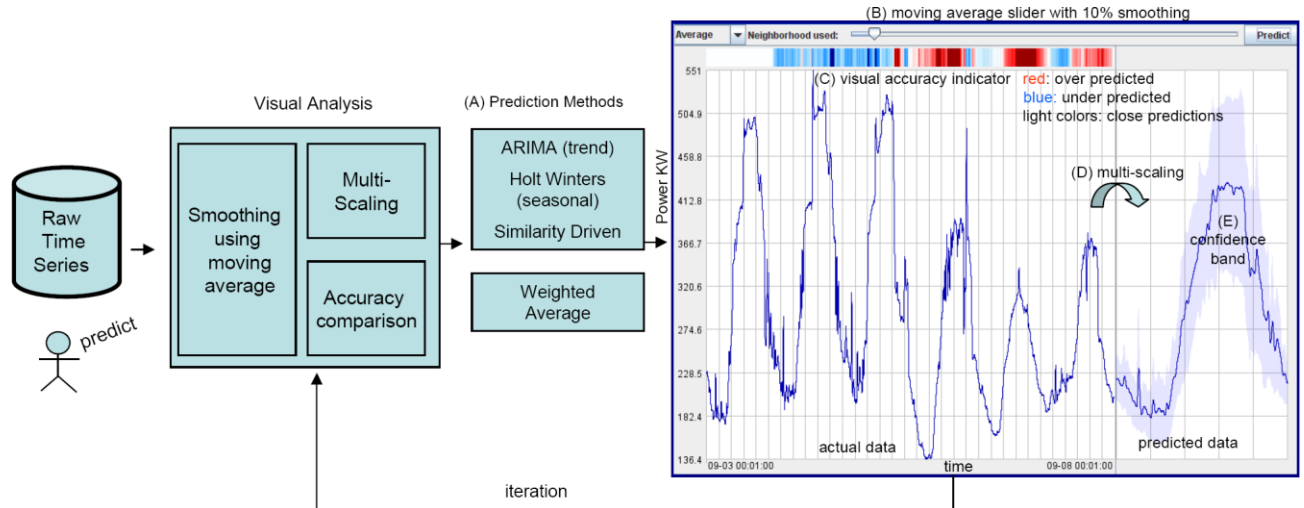Hewlett-Packard Labs and *University of Konstanz

Figure 1.   A pipeline for visual prediction of next day's power consumption KW from one week's data in a data center

ABSTRACT

Many well-known time series prediction methods have been used daily by analysts making decisions. To reach a good prediction, we introduce several new visual analysis techniques of smoothing, multi-scaling, and weighted average with the involvement of human expert knowledge. We combine them into a well-fitted method to perform prediction. We have applied this approach with success to predict resource consumption in data centers.

1   INTRODUCTION

The motivation for the analysis of a time series is the desire to predict the likelhood of the future based on historical data. Data center administrators want to predict the next day's power consumption from last month's data. A supply chain company needs to predict the number of products they must store in a warehouse this month from last year sales.

Classic statistical methods for time series prediction are model-based (e.g., ARIMA and Holt Winters [1], G-TSFE [2]) and others (smoothing, trend, similarity [3]). Each method has its own characteristics and applications. Holt Winters is used for seasonal data, while ARIMA is used for non-seasonal data. Most methods are heavily dependent on the applications and require expertise to drive the prediction process.

2   COMBINING NEW IDEAS AND EXISTING METHODS

The goal of our work is to combine existing prediction methods into a suite of visual analysis techniques to allow the user to steer the prediction process and produce better prediction results. We use a multi-scale visualization and a visual accuracy indicator in combination with standard preprocessing techniques such as smoothing with moving averages to remove noise from the time series before applying the prediction methods. We also include a technique based on weighted averages for predicting one full day's data from a monthly time series.

Figure 1 illustrates the overall process pipeline. (A) prediction methods: ARIMA, Holt Winters, Weighted Average, etc. (B)

moving average slider, (C) visual accuracy indicator, (D) multi-scaling, and (E) confidence band for use in the prediction process.

For visual prediction of large volume of data over time, we use a cell-based high-resolution time series [4] described in section 3.2 in addition to line charts to avoid overplotting.

2.1   Multi-Scaling

Multi-scaling is a visual analysis technique for the user to adjust the scale between the actual and predicted data as shown in Figure 1(D). For better visibility of the predicted data, users often assign a small portion of the screen to display the history data in order to have enough space to show the prediction results. This is essential since the predicted time frame (one day) is usually smaller than the history time frame (a week or a month).

2.2   Prediction Visual Accuracy Indicator

An important component of our system is the *visual accuracy indicator* shown in Figure 1C. This component is used to show the quality obtained by the prediction method for all time frames in the past. To determine the quality values we calculate the prediction for each time step in the past based on each measured value occurring before that time step without knowing the actual value. We normalize the differences between the predicted and actual values using standard deviation. Then, we map these values to a color map. Dark colors indicate larger differences; light colors indicate smaller differences. The red/blue lines indicate that the actual values are higher/lower respectively than the predicated value. Figure 1 first shows slightly under-predicted values (light blue lines), followed by over-predicted values (many red lines).

2.3   Smoothing with Moving Average

In our experiments with the resource consumption in data centers, in many cases the prediction methods do not provide sufficiently good results. One reason is that the data is usually very noisy. Smoothing can help to reduce the negative effects of noise on the prediction. The moving averages can be computed based on a varying time interval and therefore, we employ a moving
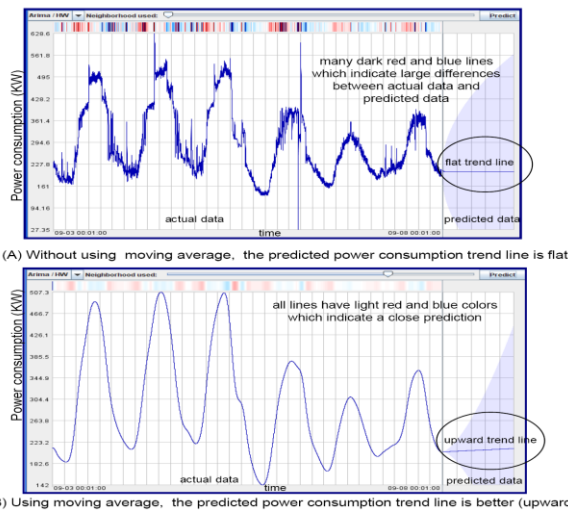
Figure 2. Apply *smoothing with moving average* to ARIMA model to predict next day's trend line on data center power consumptions

Figure 3. Two Predictions of Next Day's Chiller Utilization

average slider (Figure 1B) for users to select the time interval.

Figure 2 illustrates the effectiveness of using moving average in prediction. Figure 2B shows a better trend line (upward) than Figure 2A (straight line) using the ARIMA model on the original data.

2.4   Weighted Average Prediction

From our experience with existing prediction methods, we found that very few real-world time series fit the characteristics assumed by the different models. What is really needed is a method that combines seasonal and non-seasonal aspects and allow a prediction of longer periods such as a full day. One way of obtaining this is to use a weighted average of previous days over multiple weeks. This allows a useful long term prediction as can be seen in Figure 3.

3   APPLICATIONS & COMPARISONS

The above methods have been applied with success to real-world data center thermal prediction for administrators to plan for next day's power consumption and the percentage of chiller utilization.

3.1   Visual Prediction Line Charts

Common questions from data center administrators are: what chiller utilization can I expect tomorrow based on weekly or monthly chiller data? Which prediction methods show the better answer?

In comparing Figure 3A (Holt Winters) and Figure 3B (Weighted Average), the Weighted Average Prediction clearly shows a better prediction with more confidence than Holt Winters (more light colored lines and narrower confidence band).

3.2   Visual High Resolution Time Series

Time series used for prediction are often very long. Analysts usually use one month's historical thermal data to predict the next day's power and chiller utilization. Because of limited screen size, we use a *high-resolution cell-based* time series displayed in real-time. In Figure 4, cells (pixels) of each time series are arranged from bottom to top and left to right according to time. The size of the cells automatically scales down as more cells are displayed. The color of a cell is a one-minute measurement value for each time interval.

Analysts can simultaneously perform predictions on the chiller and power utilizations using the Holt Winters or Weighted Average Methods because both usage patterns are seasonal. The prediction result is shown on day 2008-09-09 in the last row of Figure 4. The chiller and power is going higher during the day (hours 10 to 17) and early evening (more red and orange), than
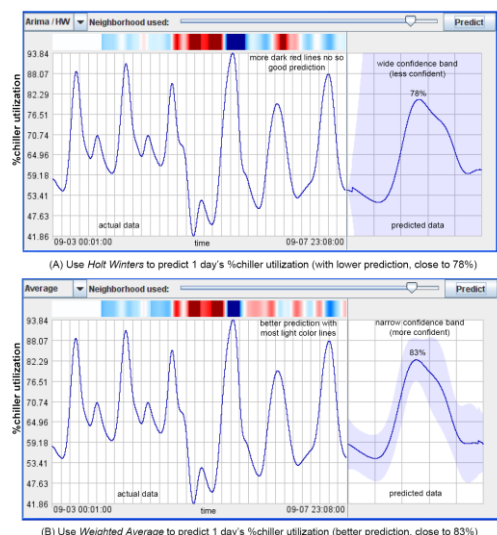
early morning and late evening (most green and yellow). This result helps the administrators plan cooling system utilization for the next day.
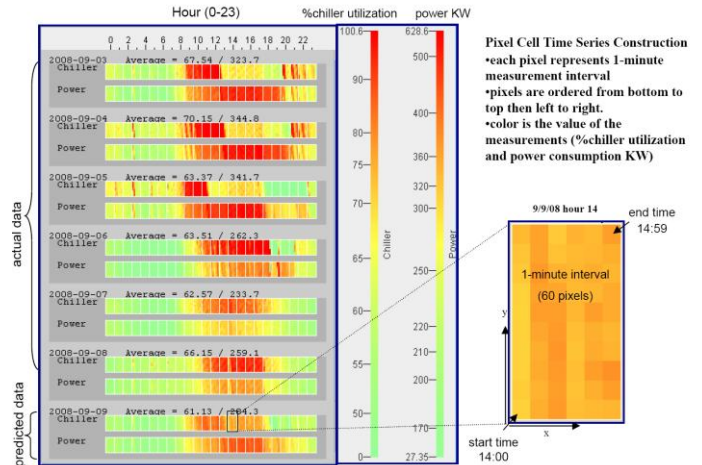

Figure 4. Multi-time series predictions on data center resource consumptions (%chiller utilization and power KW) using Holt Winters Model

4   CONCLUSION

In this poster, we presented a visual analysis of time-series prediction framework which combines existing methods and interactive visualizations for real-world data center resource consumptions. Our weighted average driven prediction method plays an important role in prediction of daily system usage. Our future work will analyze local trends and add visual indicators for high-resolution time series.

REFERENCES

[1] ARIMA and Holt Winters prediction models are described in the Chatfield, C. The analysis of time series: An Introduction, 6th ed., CRC Press, Boca Raton, USA. 2004.

[2] TSFE Time Series Automated Forecasting Engine. Castellanos, M, Norman Salazar, N., Fabio Casati, F., Dayal U. Shan M. Int. J. Computation Science and Engineering, 2000.

[3] Buono, P., et. al. Similarity-Based Forecasting with Simultaneous Previews: A River Plot Interface for Time Series Forecasting. The 11th International Conference on Information Visualization. 2007.

[4] Hao M., Dayal U., Keim D. A., Schreck T. Multi-Resolution Techniques for Visual Exploration of Large Time-Series Data. Proceedings: IEEE VGTC Symposium on Visualization, 2007.