Applied Soft Computing 9 (2009) 658–666

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A fuzzy back propagation network ensemble with example classification for lot output time prediction in a wafer fab

Toly Chen [a,*], Yu-Cheng Lin [b]

[a] Department of Industrial Engineering and System Management, Feng Chia University, Taichung City 407, Taiwan, ROC
[b] Department of Industrial Engineering and Management, Overseas Chinese Institute of Technology, Taichung City 407, Taiwan, ROC

ARTICLE INFO

Article history:
Received 14 November 2005
Received in revised form 13 September 2007
Accepted 15 April 2008
Available online 5 October 2008

Keywords:
Fuzzy back propagation network
Example classification
Output time prediction
Wafer fab

ABSTRACT

Lot output time prediction is a critical task to a wafer fabrication plant (wafer fab). To further enhance the accuracy of wafer lot output time prediction, the concept of clustering is applied to Chen's fuzzy back propagation network (FBPN) approach in this study by pre-classifying wafer lots before predicting their output times with several FBPNs that have the same topology. Each wafer lot category has a corresponding FBPN that is applied to predict the output times of all lots belonging to the category. In choosing the learning examples of each category, whether a wafer lot can be unambiguously classified or not and the accuracy of predicting the output time of the lot are simultaneously taken into account. To validate the effectiveness of the proposed methodology and to make comparison with some existing approaches, the actual data in a wafer fab were collected. According to experimental results, the prediction accuracy of the proposed methodology was significantly better than those of some existing approaches in most cases by achieving a 19–52% (and an average of 38%) reduction in the root-mean-square-error (RMSE). On the other hand, compared with the fuzzy *c*-means (FCM)-BPN-ensemble approach, the performance of the proposed methodology in the efficiency respect was indeed improved.

1. Introduction

Predicting job completion time is a critical task to various types of systems, e.g. computer systems [17], production systems [3,8], network systems [21], etc. Among them, a wafer fabrication plant (wafer fab) is a very complicated production system. A job in a wafer fab is called a wafer lot. After completing the fabrication on a wafer lot, the lot is outputted from the wafer fab. Predicting the completion/output time of a wafer lot is an even more important task to the wafer fab. After the output time of each lot in the wafer fab is accurately predicted, several managerial goals can be simultaneously achieved [8]. Predicting the output time of a wafer lot is equivalent to estimating the cycle (flow) time of the lot, because the former can be easily derived by adding the release time (a constant) to the latter. Chen [8] classified the major approaches commonly applied to predicting the output/cycle time of a wafer lot into six categories: multiple-factor linear combination (MFLC), production simulation (PS), back propagation networks (BPN), case based reasoning (CBR), fuzzy modeling methods, and hybrid approaches.

Among the six approaches, MFLC is the easiest, quickest, and most prevalent in practical applications. The major disadvantage of MFLC is the lack of forecasting accuracy [8]. Conversely, huge amount of data and lengthy simulation time are two disadvantages of PS. At first, it might take several weeks to more than 1 month to construct a full-scale simulation model for a wafer fab. Subsequently, each simulation replication usually takes a lot of minutes to several hours, and hundreds or even thousands of replications are often necessary to sufficiently consider all uncertain events. Eventually, it might take a couple of days to complete a simulation run. If any of the production conditions in the simulation model deviates from the assumed value, then another simulation run has to be executed. Nevertheless, PS is the most accurate lot output time prediction approach if the related databases are continuously updated to maintain enough validity, and often serves as a benchmark for evaluating the effectiveness/accuracy of another method. PS also tends to be preferred because it allows for computational experiments and subsequent analyses without any actual execution [3]. Considering both effectiveness and efficiency, Chang and Hsieh [2] and Chang et al. [4] both forecasted the output/cycle time of a lot in a wafer fab with a BPN having a single hidden layer. Compared with MFLC approaches, the average prediction accuracy measured in terms of root mean squared error (RMSE) was considerably improved with these BPNs. For example, an improvement of about

* Corresponding author.
E-mail address: tolychen@ms37.hinet.net (T. Chen).

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2008.04.018

40% in RMSE was achieved in Chang et al. [4]. On the other hand, much less time and fewer data are required to generate an output time forecast with a BPN than with PS. More recently, Chang et al. [3] proposed a k-nearest-neighbors based case-based reasoning (CBR) approach with dynamic factor weights and a nonlinear similarity function for due-date assignment in a wafer fab, in which the weights of factors (the cycle times of the previous cases/lots) are proportional to the similarities of the new lot with the previous cases. Chang et al.'s CBR approach outperformed the BPN approach in forecasting accuracy. In one case, the advantage was up to 27%. Chang et al. [4] modified the first step (i.e. partitioning the range of each input variable into several fuzzy intervals) of the fuzzy modeling method proposed by Wang and Mendel [26], the so-called WM method, with a simple genetic algorithm (GA) and proposed the so-called evolving fuzzy rule (EFR) approach to predict the cycle time of a lot in a wafer fab. In this way, the weakness of the WM method in selecting the number of partitions for each variable arbitrarily can be overcome. The EFR approach outperformed CBR and BPN in forecasting accuracy. Chen [8] constructed a fuzzy BPN (FBPN) that incorporated expert opinions in forming inputs to the FBPN. Chen's FBPN was a hybrid approach (fuzzy modeling and BPN) and surpassed crisp BPN approach especially in the efficiency respect. The knowledge contained in either a BPN or a CBR model is implicit and could not be easily understood by human. To enable the generation of explicit knowledge, Chang and Liao [5] constructed fuzzy rule bases with the aid of a self-organizing map (SOM) and GA. The SOM was first used to classify the data. After classification, the WM approach was then applied to extract fuzzy rules for each cluster. Later on, the fuzzy rule base could be applied to predict the output time of a new lot. Chang and Liao's SOM-WM approach was also a hybrid approach, and surpassed WM, EFR, and BPN in forecasting accuracy. As a summary, a trade-off table for selecting suitable lot output time prediction approaches refers to Chen [8].

One recent trend in this field is pre-classifying a wafer lot before estimating the cycle time of the lot, e.g. the SOM-WM approach proposed by Chang and Liao. In addition, Chen [10] constructed a k-means (kM)–FBPN system, in which a wafer lot was pre-classified by kM before predicting the cycle time of the lot by a FBPN. For embodying the uncertainty of wafer lot classification, Chen [12] used fuzzy c-means (FCM) instead. Chen [9] proposed a hybrid SOM and BPN approach for the same purpose. The results of these studies showed that pre-classifying wafer lots was beneficial to forecasting accuracy.

However, most pre-classification algorithms applied in this field could not absolutely classify wafer lots. Besides, whether the pre-classification algorithm combined with the subsequent forecasting approach was suitable for the data was questionable. To tackle these problems, Chen [11] incorporated FCM and a BPN ensemble in predicting the output time of a wafer lot, in which the BPNs of all categories were applied to predict the cycle time of a new wafer lot, and then the results were aggregated with another BPN. In this way, the error caused by mis-classification could be reduced. Though the prediction accuracy of the FCM-BPN-ensemble approach was good, a lengthy time was required to train and test the BPN-ensemble.

To solve these problems and to further enhance the effectiveness or efficiency of wafer lot output time prediction, the following tasks are done in this study:

(1) The FCM-BPN-ensemble approach is modified in this study by replacing the BPN-ensemble with a FBPN-ensemble. According to the results of some previous studies (e.g. [7,8]), a FBPN often converges faster than the crisp BPN does. In this way, the efficiency of wafer lot output time prediction can be improved.
(2) In choosing the learning examples of each category, whether a wafer lot can be unambiguously classified or not and the accuracy of predicting the output time of the lot are simultaneously taken into account. This way guarantees that if a new lot is unambiguously classified into a category, then its output time can be accurately predicted with the FBPN of the category. Otherwise, the wafer lot is not adopted as an example by any category, and the FBPNs of all categories (i.e. the FBPN-ensemble) is applied to predict the output time of the lot.
(3) To aggregate the outputs from the component FBPNs, another FBPN is constructed, which is also expected to be able to quicken the speed of aggregation.

The remainder of this paper is organized as follows. Section 2 introduces the FCM and FBPN-ensemble approach. To evaluate the effectiveness of the proposed methodology, the historical data in a wafer fab were collected, and then the proposed methodology and some existing approaches were applied to the collected data in Section 3. Based on analysis results, some points are made in Section 4. Finally, the concluding remarks and some directions for future research are given in Section 5.

2. Methodology

Variables that are used in the proposed methodology are defined:

(1) R_n: the release time of lot n.
(2) U_n: the average fab utilization at R_n. $\tilde{U}_n$ is derived by multiplying the importance of U_n that is expressed with a fuzzy value to U_n.
(3) Q_n: the total queue length on the processing route of lot n at R_n. $\tilde{Q}_n$ is derived by multiplying the importance of Q_n that is expressed with a fuzzy value to Q_n.
(4) BQ_n: the total queue length before bottlenecks at R_n. $B\tilde{Q}_n$ is derived by multiplying the importance of BQ_n that is expressed with a fuzzy value to BQ_n.
(5) FQ_n: the total queue length in the whole fab at R_n. $F\tilde{Q}$ is derived by multiplying the importance of FQ_n that is expressed with a fuzzy value FQ_n.
(6) WIP_n: the fab work-in-progress (WIP) at R_n. $\widetilde{WIP}_n$ is derived by multiplying the importance of WIP_n that is expressed with a fuzzy value to WIP_n.
(7) $D_n^{(i)}$: the delay of the i-th recently completed lot, i = 1–3. $\tilde{D}_n^{(i)}$ is derived by multiplying the importance of $D_n^{(i)}$ that is expressed with a fuzzy value to $D_n^{(i)}$.
(8) CT_n: the cycle time of lot n.
(9) $(+)$, $(-)$, $(\times)$: fuzzy addition, subtraction, and multiplication, respectively.

There are eleven stages in the proposed methodology (see Fig. 1):

(1) Collect the historical data of old wafer lots.
(2) Classify the old wafer lots using FCM, and record the membership of each lot belonging to every category (indicated with μ_{n1}).
(3) Predict the cycle time of each wafer lot with the FBPN of every category, and record the prediction error.
(4) For each category, construct a triangular fuzzy number (TFN) to convert the prediction accuracy into a membership function value (indicated with μ_{n2}) (see Fig. 2).
(5) If $\mu_{n1}{\cdot}\mu_{n2} < h$ (a threshold between 0 and 1), then the lot is excluded from the examples for training and testing the FBPN of the category.