Computers and Chemical Engineering 79 (2015) 113–134

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Planning and scheduling of steel plates production. Part I: Estimation of production times via hybrid Bayesian networks for large domain of discrete variables

Junichi Mori, Vladimir Mahalec*

Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4L7

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form 4 February 2015
Accepted 8 February 2015
Available online 5 March 2015

Keywords:
Estimation of production time
Hybrid Bayesian network
Structure learning
Large domain of discrete variables
Decision tree
Steel plate production

ABSTRACT

Knowledge of the production loads and production times is an essential ingredient for making successful production plans and schedules. In steel production, the production loads and the production times are impacted by many uncertainties, which necessitates their prediction via stochastic models. In order to avoid having separate prediction models for planning and for scheduling, it is helpful to develop a single prediction model that allows us to predict both production loads and production times. In this work, Bayesian network models are employed to predict the probability distributions of these variables. First, network structure is identified by maximizing the Bayesian scores that include the likelihood and model complexity. In order to handle large domain of discrete variables, a novel decision-tree structured conditional probability table based Bayesian inference algorithm is developed. We present results for real-world steel production data and show that the proposed models can accurately predict the probability distributions.

1. Introduction

Accurate estimation of the production loads and the total production times in manufacturing processes is crucial for optimal operations of real world industrial systems. In this paper, a production load represents the number of times that a product is processed in the corresponding process unit and a production time is defined as the length of time from production start to completion. Production planning and scheduling for short, medium and long term time horizons employ various optimization models and algorithms that require accurate knowledge of production loads and total production times from information at each of the processing steps. The approaches to predict the production loads and total production times can be either based on mechanistic model or can employ data-driven techniques. Model-based prediction methods may be applied only if accurate mechanistic models of the processes can be developed. First principal models require in-depth knowledge of the processes and still cannot take into consideration all uncertainties that exist in the processes. Therefore, mechanistic models may not work well for prediction of production loads and production times of the real-world industrial processes. On the other hand, data-driven approaches do not require in-depth process knowledge and some advanced techniques can deal with the process uncertainties.

The most straightforward approach is to use the classical statistical models (e.g. regression models) that estimate the values of new production loads and a production time from the past values in the historical process data. The relationships between the targeted variables and other relevant variables are used to compute the statistical model that can predict the production loads and production times. However, such models are too simple to predict the nonlinear behavior and estimate the system uncertainties. An alternative simple method is to compute the average of these values per each production group that has similar production properties and then utilize the average values of each production group as the prediction (Ashayeria et al., 2006). In this case the prediction accuracy significantly depends on the rules that govern creation of production groups and it may be challenging to find the appropriate rules from process knowledge only. To overcome this limitation, supervised classification techniques such as artificial neural networks (ANN), support vector machine, Fisher discriminant analysis and K-nearest neighbors (KNN) may be useful to design the rules to make production groups from historical process data. However, even though we can accurately classify the historical process data into an appropriate number of production groups, these methods do not consider model uncertainties and cannot handle missing

* Corresponding author. Tel.: +1 905 525 9140.
E-mail address: mahalec@mcmaster.ca (V. Mahalec).

http://dx.doi.org/10.1016/j.compchemeng.2015.02.005
0098-1354/

values and unobserved variables. In addition, we are typically forced to have multiple specific models tailored for specific purposes, e.g. for production planning or for scheduling, which causes model maintenance issues and lack of model consistency.

Bayesian network (BN) models offer advantages of having a single prediction model for predicting the production loads and total process times in planning and scheduling. Bayesian networks are also called *directed graphical models* where the links of the graphs represent direct dependence among the variables and are described by arrows between links (Pearl, 1988; Bishop, 2006). Bayesian network models are popular for representing conditional independencies among random variables under system uncertainty. They are popular in the machine learning communities and have been applied to various fields including medical diagnostics, speech recognition, gene modeling, cancer classification, target tracking, sensor validation, and reliability analysis.

The most common representations of conditional probability distributions (CPDs) at each node in BNs are conditional probability tables (CPTs), which specify marginal probability distributions for each combination of values of its discrete parent nodes. Since the real industrial plant data often include discrete variables which have large discrete domains, the number of parameters becomes too large to represent the relationships by the CPTs. In order to reduce the number of parameters, context-specific independence representations are useful to describe the CPTs (Boutilier et al., 1996). Efficient inference algorithm that exploits context-specified independence (Poole and Zhang, 2003) and the learning methods for identification of parameters of context-specific independence (Friedman and Goldszmidt, 1996; Chickering et al., 1997) have been developed. The restriction of these methods is that all discrete values must be already grouped at an appropriate level of domain size since learning structured CPTs is NP-hard. However, discrete process variables typically have large domains and the task of identifying a reasonable set of groups that distinguish well the values of discrete variables requires in-depth process knowledge. To overcome this limitation, *attribute – value hierarchies* (AVHs) that capture meaningful groupings of values in a particular domain are integrated with the tree-structured CPTs (DesJardins and Rathod, 2008). Such approach is not applicable in general process systems, since some discrete process variables do not contain hierarchal structures and thus AVHs cannot capture the useful abstracts of values in that domain. In addition, this model cannot handle the continuous variables without discretizing them. Furthermore, the authors do not describe how to apply AVH-derived CPTs to Bayesian inference. Therefore, this method has difficulty predicting probability distributions of production loads and total process time from observed process variables in the real-world industrial processes. Efficient alternative inference methods in Bayesian Networks containing CPTs that are represented as decision trees have been developed (Sharma and Poole, 2003). The inference algorithm is based on variable elimination (VE) algorithm. However, because the computational complexity of the exact inference such as VE grows exponentially with the size of the network, this method may not be appropriate for Bayesian networks for large scale industrial data sets. In addition, the method does not deal with application of the decision-tree structured CPTs to hybrid Bayesian networks, where both discrete and continuous variables appear simultaneously.

In hybrid Bayesian networks, the most commonly used model that allows exact inference is the conditional linear Gaussian (CLG) model (Lauritzen, 1992; Lauritzen and Jensen, 2001). However, the proposed network model does not allow discrete variables to have continuous parents. To overcome this limitation, the CPDs of these nodes are typically modeled as softmax function, but there is no exact inference algorithm. Although an approximate inference via Monte Carlo method has been proposed (Koller et al., 1999), the convergence can be quite slow in Bayesian Networks with large domain of discrete variables. Another approach is to discretize all continuous variables in a network and treat them as if they are discrete (Kozlov and Koller, 1997). Nevertheless, it is typically impossible to discretize the continuous variables as finely as needed to obtain reasonable solutions and the discretization leads to a trade-off between accuracy of the approximation and cost of computation. As another alternative, the mixture of truncated exponential (MTE) model has been introduced to handle the hybrid Bayesian networks (Moral et al., 2001; Rumi and Salmeron, 2007; Cobb et al., 2007). MTE models approximate arbitrary probability distribution functions (PDFs) using exponential terms and allow implementation of inference in hybrid Bayesian networks. The main advantage of this method is that standard propagation algorithms can be used. However, since the number of regression coefficients in exponential functions linearly grows with the domain size of discrete variables, MTE model may not work well for the large Bayesian networks that are required to represent the industrial processes.

Production planning and scheduling in the steel industry are recognized as challenging problems. In particular, the steel plate production is one of the most complicated processes because steel plates are high-variety low-volume products manufactured on order and they are used in many different applications. Although there have been several studies on scheduling and planning problems in steel production, such as continuous casting (Tang et al., 2000; Santos et al., 2003), smelting process (Harjunkoski and Grossmann, 2001) and batch annealing (Moon and Hrymak, 1999), few studies have dealt with steel plate production scheduling. Steel rolling processes manufacture various size of plates from a wide range of materials. Then, at the finishing and inspection processes, malfunctions occurred in the upstream processes (e.g. smelting processes) are repaired, and additional treatments such as heat treatment and primer coating are applied such that the plates satisfy the intended application needs and satisfy the demanded properties. In order to obtain successful plans and schedules for steel plate production, it is necessary to determine the production starting times that meet both the customer shipping deadlines and the production capacity. This requires prediction models that can accurately predict the production loads of finishing and inspection lines and the total process time. However, due to the complexity and uncertainties that exist in the steel production processes, it is difficult to build the precise prediction models. These difficulties have been discussed in the literature (Nishioka et al., 2012).

In our work, in order to handle the complicated interaction among process variables and uncertainties, Bayesian networks are employed for predicting of the production loads and prediction of the total production time. Since the steel production data have large domain of discrete variables, their CPDs are described by tree structured CPTs. In order to compute the tree structured CPTs, we use decision trees algorithm (Breiman et al., 1984) that is able to group the discrete values to capture important distinctions of continuous or discrete variables. If Bayesian networks include continuous parent nodes with a discrete child node, the corresponding continuous variables can be discretized as finely as needed, because the domain size of discretized variable does not increase the number of parameters in intermediate factors due to decision-tree structured CPTs. Since the classification algorithms are typically greedy ones, the computational task for learning the structured CPTs is not expensive. Then, the intermediate factors can be described compactly using a simple parametric representation called the canonical table representation.

As for Bayesian network structure, if the cause–effect relationship is clearly identified from process knowledge, knowledge based network identification approach is well-suited. Such identification of cause–effect relationships may require in-depth knowledge of the processes to characterize the complex physical, chemical and