Journal of Process Control 35 (2015) 169–177

Contents lists available at ScienceDirect

Journal of Process Control

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

A Two-stage Clustered Multi-Task Learning method for operational optimization in Chemical Mechanical Polishing

Yunqiang Duan[a], Min Liu[a,b,*], Mingyu Dong[a], Cheng Wu[a]

[a] Department of Automation, Tsinghua University, Beijing 100084, PR China
[b] Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, PR China

ARTICLE INFO

Article history:
Received 15 December 2014
Received in revised form 14 April 2015
Accepted 3 June 2015
Available online 25 September 2015

Keywords:
Operational optimization
Clustered Multi-Task Learning
Chemical Mechanical Polishing
Small sample size
Modelling

ABSTRACT

Operational optimization of Chemical Mechanical Polishing, which sets the proper polishing time, is very important for improving the production efficiency of semiconductor manufacturing processes. However, usual operational optimization methods based on Run-to-Run strategies have not been suitable for the mixed-product processing mode of CMP. Also, under the mode, it is very difficult to model the polishing time due to the insufficient number of the corresponding samples. In this paper, a Two-stage Clustered Multi-Task Learning method is proposed for the above modelling problem with small sample size, in which the proposed Probability-based Task Clustering algorithm first groups similar products so that their corresponding samples can be used for modelling simultaneously. After this, in each cluster, the proposed Shared Multi-Task Learning (SMTL) algorithm obtains the corresponding model for each kind of products cooperatively, in which the parameter vector of each model is the sum of two parts – the shared part and the private part. In each cluster, the shared part represents the common characteristics of all products and the private part represents the particular characteristics of each kind of products. Also, in SMTL, the two parts can be obtained after a non-smooth convex optimization problem is constructed and solved through the Accelerated Proximal Method. The results of numerical simulations on a practical industrial data set and the other two data sets demonstrate the effectiveness of the proposed algorithms. The proposed algorithms can also be used in other problems such as the modelling problems of key indexes of urban development and operation.

1. Introduction

Due to the huge demand of electronic products, great efforts have been made to improve the production efficiency in the semiconductor manufacturing field. For this purpose, manufacturers must optimize the process of Chemical Mechanical Polishing (CMP) since it has a strong impact on the production efficiency of the whole semiconductor manufacturing process. More specifically, CMP is a process of smoothing the surface of a wafer with a combination of mechanical and chemical forces [1] in a period of time called polishing time. During the manufacturing process, once a defective wafer processed by CMP is found, the whole lot, which consists of many wafers, has to be returned to the former process. Obviously, the above rework procedure is time-consuming. Thus, for cutting down the rework rate, manufacturers must implement the operational optimization of CMP so as to obtain the suitable polishing time.

In recent years, many operational optimization and control methods are designed for CMP. Based on Run-to-Run (RtR) strategies, some methods assume that the same kind of products are processed on one machine successively [1–4]. Other methods assume that different kinds of products are processed on one machine in a fixed pattern [5–7]. However, for the demands of the actual production, the mixed-product processing mode is often applied. In this mode, one machine may process different kinds of products in any pattern. It is obvious that the above assumptions are not consistent with the mixed-product processing mode. So the suitable polishing time cannot be obtained automatically. In consequence, many manufacturers have to determine the polishing time manually which leads to the high rework rate and the low product efficiency under the real changeable production environments.

In order to implement the operational optimization of CMP (i.e. setting the suitable polishing time) in the mixed-product processing mode, we model the polishing time for each kind of products based on the corresponding history data (samples). But

* Corresponding author.
E-mail address: lium@tsinghua.edu.cn (M. Liu).

1 http://en.wikipedia.org/wiki/Chemical-mechanical_planarization

http://dx.doi.org/10.1016/j.jprocont.2015.06.005
0959-1524/

in the real situation, for most kinds of products, the number of samples are small and it is difficult to obtain the accurate models through the traditional modelling methods. Fortunately, there are a large number of products and the total number of samples of all kinds of products are large. Accordingly, in this paper, we intend to group the corresponding samples of similar products (group similar products for short) so that their corresponding samples can be used for modelling simultaneously. This can be achieved through Multi-Task Learning (MTL), specifically Clustered Multi-Task Learning (CMTL), in which one task corresponds to the modelling problem for one kind of products.

As discussed in Section 3, the existing CMTL algorithms generally achieve task clustering and model learning at the same time. Hence, these two parts have a serious interaction and it is difficult to optimize and use them separately. In addition, they do not consider the situation of small sample size and some priori knowledge about task clustering is also difficult to be utilized, as requested in the operational optimization of CMP. So in this paper, we propose a novel CMTL algorithm called Two-stage Clustered Multi-Task Learning (TwCMTL) for the operational optimization of CMP.

In the first stage, a Probability-based Task Clustering (PTC) algorithm, which is based on the probability divergence and the Affinity Propagation (AP) [8], is developed for grouping similar products. In this way, the samples of similar products can be used for modelling simultaneously. Also, a strategy based on the sample size information of all products is designed to select cluster centers from products with larger sample size so that each cluster has sufficient samples.

In the second stage, a novel MTL algorithm called Shared Multi-Task Learning (SMTL) is applied in each cluster to obtain the corresponding model for each kind of products cooperatively. The parameter vector (including several parameters) of each model is the sum of the shared part and the private part. In each cluster, there is one shared part and each model has its own private part. The shared part stands for the common characteristics of all products and the private part represents the particular characteristics of each kind of products. In this way, the similarities among products can be represented by the values of these two parts. When the products are similar to each other, both the number and the value of the non-zero private parts should be small. Especially, when the sample sizes are small, SMTL can be adjusted to focus on the shared part since in this case the shared part is more reliable. Furthermore, the proposed SMTL algorithm learns the above two parts simultaneously after a constructed non-smooth convex optimization problem is solved through the Accelerated Proximal Method (APM) efficiently. In addition, since the small sample problem is common, the proposed algorithms can also be used in other practical modelling problems such as the modelling problems of key indexes of urban development and operation.

The main contributions of this study include:

1. For the operational optimization of CMP, we propose a two-stage CMTL algorithm called TwCMTL to model the polishing time for each kind of products even when the number of the corresponding samples is small. TwCMTL groups similar products and learns one model for each kind of products.
2. In order to group similar products, we propose a PTC algorithm based on the probability divergence and the AP algorithm. It also utilizes the sample size information of all products in the process of clustering.
3. We design a novel SMTL algorithm and use it as the model learning algorithm in TwCMTL. SMTL learns the shared part and the private part of model parameters. Particularly, the APM is employed to solve the optimization problem in SMTL efficiently.

Table 1
Notations

Notations	Explanations
$\aleph_m$	Set $\{1, \ldots, m\}$
$\Re^n$	n-Dimensional Euclidean space
a, $\boldsymbol{a}$, A	Scale a, vector $\boldsymbol{a}$, matrix a
x_i	The ith element of vector $\boldsymbol{x}$
x_{ij}	The element of matrix X in row i and column j
$\boldsymbol{x}_{(i)}$	The ith row of matrix X
$\boldsymbol{x}^{(i)}$	The ith column of matrix X
$\\|\boldsymbol{x}\\|_p = (\Sigma_i x_i^p)^{\frac{1}{p}}$	The p-norm of vector $\boldsymbol{x}$
$\\|X\\|_{p,q} = (\Sigma_i((\Sigma_j x_{ij}^p)^{\frac{1}{p}})^q)^{\frac{1}{q}}$	The $l_{p,q}$-norm of matrix X
$[a\,;b\,;c]$	The column vector composed by a, b and c
$\\|X\\|_F$	The Frobenius norm of matrix X
$\|X\|$	The determinant of matrix X
$\langle *, * \rangle$	The inner product of two vectors or two matrixes
$*'$	The transposition of a vector or a matrix
$\mathcal{N}(\mu, \sigma^2)$	The Gaussian distribution with mean μ and variance σ^2
$\mathcal{U}(a, b)$	The uniform distribution between a and b

Table 2
Abbreviations

Abbreviations	Full name
CMP	Chemical Mechanical Polishing
MRR	Material Removal Rate
RtR	Run-to-Run
MTL	Multi-Task Learning
RMTL	Robust Multi-Task Learning
CMTL	Clustered Multi-Task Learning
RMTFL	Robust Multi-Task Feature Learning
kCMTL	The relaxed k-means clustering for CMTL
cCMTL	The convex formulation for CMTL
AP	Affinity Propagation
PTC	Probability-based Task Clustering
SMTL	Shared Multi-Task Learning
TwCMTL	Two-stage Clustered Multi-Task Learning
APM	Accelerated Proximal Method
MGD	Multivariate Gaussian Distribution

The structure of this paper is stated as follows. In Section 2, the modelling problem of polishing time for CMP is described. Section 3 reviews the MTL and CMTL algorithms. Next, Section 4 proposes the TwCMTL algorithm including the two-step framework for CMTL, the proposed PTC and the novel SMTL. After this, the numerical simulations are given in Section 5. Finally, the conclusion and the further works are stated in Section 6. Some important notations and abbreviations are listed in Tables 1 and 2, respectively.

2. The modelling problem of polishing time for CMP

CMP, which produces a planar surface for the subsequent processes, is an important process for semiconductor manufacturing processes. The diagram of CMP is shown in Fig. 1, in which the uneven surface of one wafer is polished by an abrasive pad and

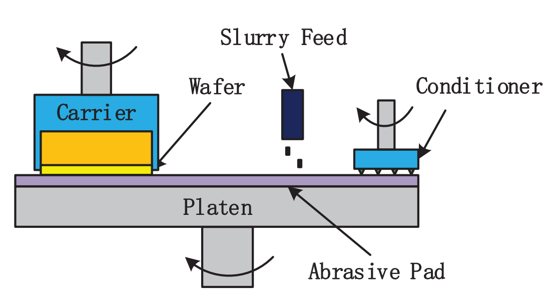

Fig. 1. Processes of CMP (side view).