**Title:** Throughput Rate Estimation for Operations in Semiconductor Fabrication

**Names of group members:** Liam Hsieh

**Group number:** 23

**Abstract:** An accurate and efficient estimate of throughput rate is crucial for capacity planning in semiconductor manufacturing, as the significant amount of equipment depreciation constantly demands improved operational planning approaches. To optimize business objectives under uncertainty using advanced optimization techniques, it is necessary to have estimates of throughput rate distributions and a systematic solution to reduce decision-making lead time. The goal of creating a systematic solution is to fully automate the process of data collection, data cleansing and parameter estimation. The project primarily focuses on the part of throughput rate estimation where key features suggested by literature are extracted from an asynchronous database behind the Manufacturing Execution System (MES), and domain experts provide labels for dataset classification based on processing behaviors as a temporary solution. Random Forest Regression is applied to predict the throughput rate and the experiment shows a satisfactory result.

**Introduction:**

In semiconductor manufacturing, a capital-intensive and process-complex industry, capacity planning is critical for optimally utilizing available production resource which typically count on mathematical programming models, and tailored algorithms are required to efficiently solve these NP-hard problems [2].

Although model formulation may vary among different cases, the essential part of this sort of models is similar, for example:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Max**  **Subject to**  The other constraints  All decision variables | Parameters   |  |  | | --- | --- | |  | demand of product *i* in time *b* | |  | throughput rate of machine *r* while executing recipe *s* for product *i* | |  | unit selling price of product *i* in time *b* | |  | available time of machine *r* in time *b* |   Variables   |  |  | | --- | --- | |  | plan quantity of product *i* in time *b* | |  | job assignment for product *i* executing recipe *s* on machine *r* in time *b* | |

This example aims at maximizing the total revenue subject to a few constraints such as capacity restriction, job assignment & plan quantity balancing, and demand fulfillment. Please note that there are more constraints to be formulated to meet the operational needs in practice.

Throughput rate, is usually estimated via observing processing time since they are reciprocals for each other, and the estimation for all recipes is a tremendous amount of work including data collection, data cleansing, and statistical inference. Moreover, the limited decision-making lead time is a great challenge since the stochastic processing time is a kind of nature in semiconductor manufacturing [3], and modern approaches also consider more uncertainty into account for risk management. Therefore, a throughput rate forecast model is required to efficiently feed optimization models under different scenarios.

**Related work:**

Cycle time, also called flow time, is a very critical factor in semiconductor manufacturing for operations management; it is the sum of processing time and non-processing time including waiting time and transportation time. Product cycle time can be seen as the sum of step flow time.

Many approaches were proposed to forecast cycle time such as (multiple) linear regression [6], queuing models, fuzzy c-means [1], and artificial neural networks [8]. Discrete event simulation is another popular approach for cycle time estimation, it is not only for performance evaluation of semiconductor wafer fabs, but also for predicting performance measures with Monte Carlo methods [9]. However, building statistically significant simulation models could be time-consuming which often dramatically increases the decision-making lead time.

Another key issue for this project is key features selection. Meidan *et al.* [4] proposed a selective naïve Bayesian classifier and narrowed the list of factors from 182 to 20 for selection of a minimal, most discriminative key-factor set for cycle time prediction. Schelthoff [5] proposed a framework for feature selection for waiting time predictions in semiconductor wafer fabs. Tirkel [7] determined 19 out of 37 features considering attributes’ redundancy for flow time forecasting using knowledge discovery in databases. Those findings and suggestions played as a great reference for this project on data collection.

**Problem formulation:**

This project applies machine learning to predict processing time (PT) and convert it into throughput rate for capacity planning models. The approach involves using a two-step process: a classification model to determine the operational category for a recipe, followed by a regression model to predict the recipe's PT. Domain expert labels based on feature analysis and time motion study are used for the classification model. Random Forest is employed as the machine learning model due to its robustness to overfitting and capability of handling feature interactions and non-linear relationships.

A picture containing text, screenshot, font, white

Description automatically generated

**Experimental Results:**

Thirteen key factors were selected for the model, and the dataset comprised 19,500 rows. The data was coded to follow security policies, and the features included PT, TYPE, ENTITY, RECIPE, LOT\_SIZE, DATE, PRE\_UTIL\_MA, POST\_UTIL\_MA, NUM\_QUAL\_TOOLS, X\_FACTOR, OOC\_MA, OOS\_MA, WAFER\_COUNT, and MAX\_WC2PM. The Random Forest model, implemented using the scikit-learn library in Python, yielded satisfactory prediction performance as demonstrated by histogram for each recipe.

A picture containing text, diagram, plan, line

Description automatically generated

The basic information for these features:

|  |  |  |
| --- | --- | --- |
| **Column name** | **Dtype** | **Remark** |
| PT | numeric | Processing time |
| TYPE | category | Behavior label |
| ENTITY | category | Machine identity |
| RECIPE | category | Recipe identity |
| LOT\_SIZE | numeric | Number of wafers per lot |
| DATE | datetime | Time stamp of transaction |
| PRE\_UTIL\_MA | numeric | Moving average of machine utilization of previous steps |
| POST\_UTIL\_MA | numeric | Moving average of machine utilization of post steps |
| NUM\_QUAL\_TOOLS | numeric | Number of qualification machines |
| X\_FACTOR | numeric | Provided by queueing models |
| OOC\_MA | numeric | Moving average of out-of-control-limit |
| OOS\_MA | numeric | Moving average of out-of-spec |
| WAFER\_COUNT | numeric | Accumulated processed wafers since last down |
| MAX\_WC2PM | numeric | Wafer-based target for preventive maintenance |

**Contributions:**

* Literature review
* ELT data pipelines, data cleansing jobs, and data warehouse
* Python scripts
* Finished report and slides deck

**Acknowledgements:**

The data source for this project was an asynchronous MES database for an Intel 300mm Fab. The author implemented the ELT data pipeline and received support and knowledge transfer from anonymous industrial engineers and process engineers for providing meaningful labels.

**References:**

1. Chen, Toly. "An efficient and effective fuzzy collaborative intelligence approach for cycle time estimation in wafer fabrication." *International Journal of Intelligent Systems* 30.6 (2015): 620-650.
2. Hsieh, Liam Y., and Chi-Bin Cheng. "Efficient due-date quoting and production scheduling for integrated circuit packaging with reentrant processes." *IEEE Transactions on Components, Packaging and Manufacturing Technology* 8.8 (2018): 1487-1495.
3. Hsieh, Liam Y., and Tsung-Ju Hsieh. "A throughput management system for semiconductor wafer fabrication facilities: Design, systems and implementation." *Processes* 6.2 (2018): 16.
4. Meidan, Yair, et al. "Cycle-time key factor identification and prediction in semiconductor manufacturing using machine learning and data mining." *IEEE transactions on semiconductor manufacturing* 24.2 (2011): 237-248.
5. Schelthoff, Kai, et al. "Feature Selection for Waiting Time Predictions in Semiconductor Wafer Fabs." *IEEE Transactions on Semiconductor Manufacturing* 35.3 (2022): 546-555.
6. Sha, D. Y., R. L. Storch, and C-H. Liu. "Development of a regression-based method with case-based tuning to solve the due date assignment problem." *International Journal of Production Research* 45.1 (2007): 65-82.
7. Tirkel, Israel. "Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases." *International Journal of Production Research* 51.18 (2013): 5536-5548.
8. Wang, Junliang, Jie Zhang, and Xiaoxi Wang. "A data driven cycle time prediction with feature selection in a semiconductor wafer fabrication system." *IEEE Transactions on Semiconductor Manufacturing* 31.1 (2018): 173-182.
9. Xu, Jie, et al. "Simulation optimization in the era of Industrial 4.0 and the Industrial Internet." *Journal of Simulation* 10.4 (2016): 310-320.