

Applying Data Mining Techniques to Address Critical Process Optimization Needs in Advanced Manufacturing

Li Zheng¹, **Chunqiu Zeng¹**, Lei Li¹, Yexi Jiang¹, Wei Xue¹, Jingxuan Li¹, Chao Shen¹, Wubai Zhou¹, Hongtai Li¹, Liang Tang¹, Tao Li¹, Bing Duan², Ming Lei² and Pengnian Wang²

¹School of Computer Science, Florida International University, Miami, FL, USA

²ChangHong COC Display Devices Co., Ltd, Mianyang, Sichuan, China

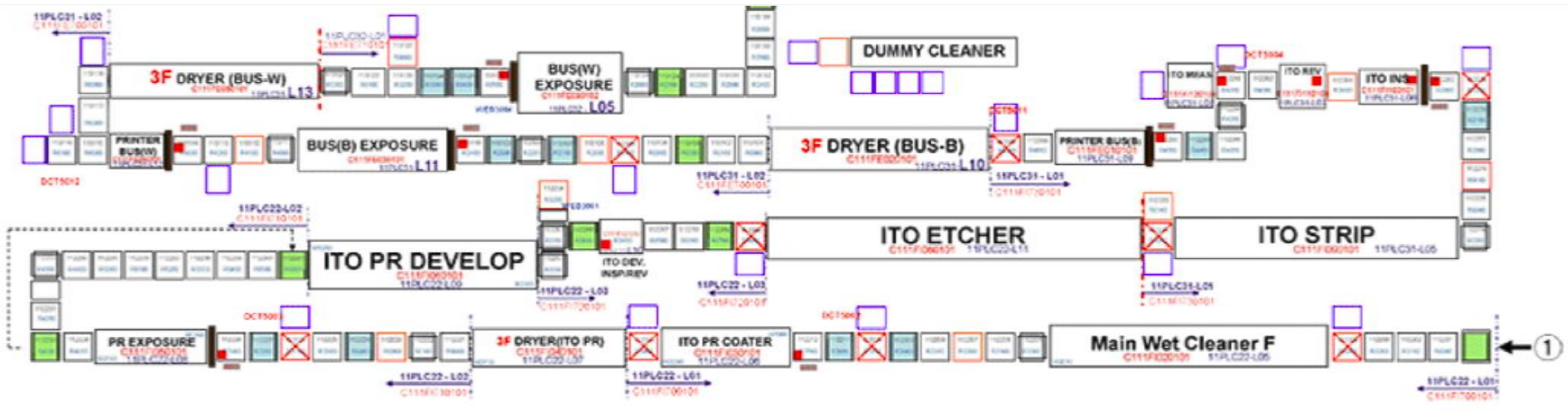
Outline

- Introduction
- Motivation
- System Overview
- Data Analysis
 - Ensemble Feature Selection
 - Regression Analysis
 - Discriminative Analysis
- System Deployment
- Conclusion

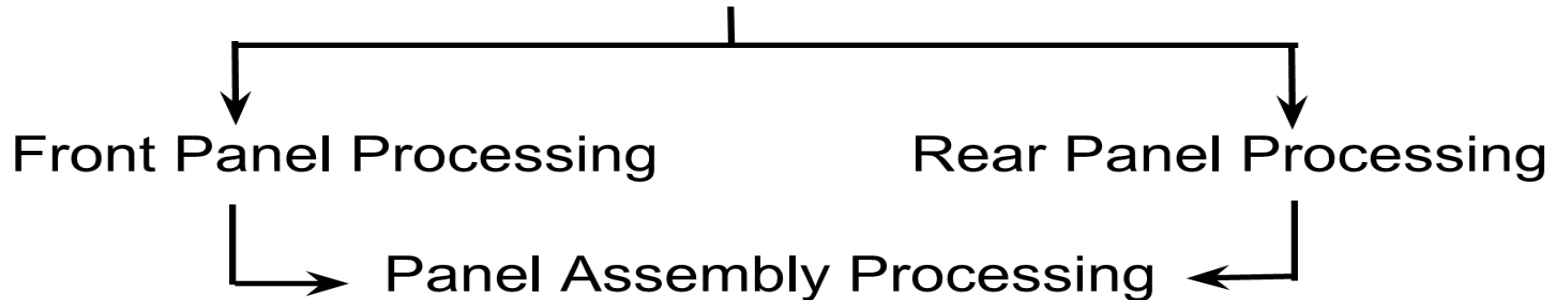
Introduction

- Advanced manufacturing refers to the manufacturing processes improved by new technologies.
 - cutting-edge materials
 - emerging capabilities enabled by physics, biology, chemistry, computer sciences, etc.
- In advanced manufacturing, **complex** and **elaborate** production processes are arranged according to the product structure.

PDP(Plasma Display Panel) Manufacturing from ChangHong



PDP Manufacturing Production Flow



3 major procedures

75 assembly routines

279 major equipments

over 10,000 parameters

6000m production line

76hr processing time

Introduction

- It generates a huge amount of production data (10G per day with 30M records)
- It contains intricate dependencies among a tremendous amount of controlling parameters(10,000 parameters).

Motivation

- In daily operation, the manufacturers are concerned with how to improve the yield rate of production.
- Manufacturers concern with such capabilities:
 - Discovering the most related parameters;
 - What are the key parameters whose values can significantly differentiate qualified products from defective products?
 - Quantifying the parameter correlation with the product quality;
 - How do the parameter value changes affect the production rate?
 - Finding novel parameter recipes to improve the production rate.
 - What are the effective parameter value combinations to ensure high yield rate?

Motivation

- Problem: application gap between manufacturers and data analysts
 - ✓ Capacity: what the data looks like?
 - ✓ Capability: How the data can be utilized?
 - ✓ Knowledge: How to perform knowledge discovery and management?
- Provide manufacturers with an integrated data analysis platform.

	Capacity	Capability	Knowledge
Manufacturers	<ul style="list-style-type: none">● huge production output● sophisticated workflow● complex supply chain	<ul style="list-style-type: none">● control yield rate● optimize production line● effective parameter setting	<ul style="list-style-type: none">● private Know-How● high dependency to experts● high cost of testing
Data Analysts	<ul style="list-style-type: none">● large number of samples● high-dimensional data● complex param dependencies	<ul style="list-style-type: none">● process optimization● feature reduction and selection● feature association analysis	<ul style="list-style-type: none">● utilize domain expertise● knowledge sharing● knowledge management
Application Gap	<ul style="list-style-type: none">● utilize customized data analysis algorithms to mine the underlying knowledge;● provide configurable task platforms to allow automatic taskflow execution;● enable efficient knowledge representation and management.		

Challenges

- Facing the enormous data with sustained growth:
 - ✓ efficiently support large-scale data analysis tasks;
 - ✓ provide prompt guidance to different routines in the workflow.

- An integrated data analysis platform is designed and implemented based on the distributed system to support high-performance analysis.

Challenges

- Facing various types of mining requirements,
 - ✓ how to effectively adapt existing algorithms for customized analysis tasks that comprehensively consider the domain characteristics?

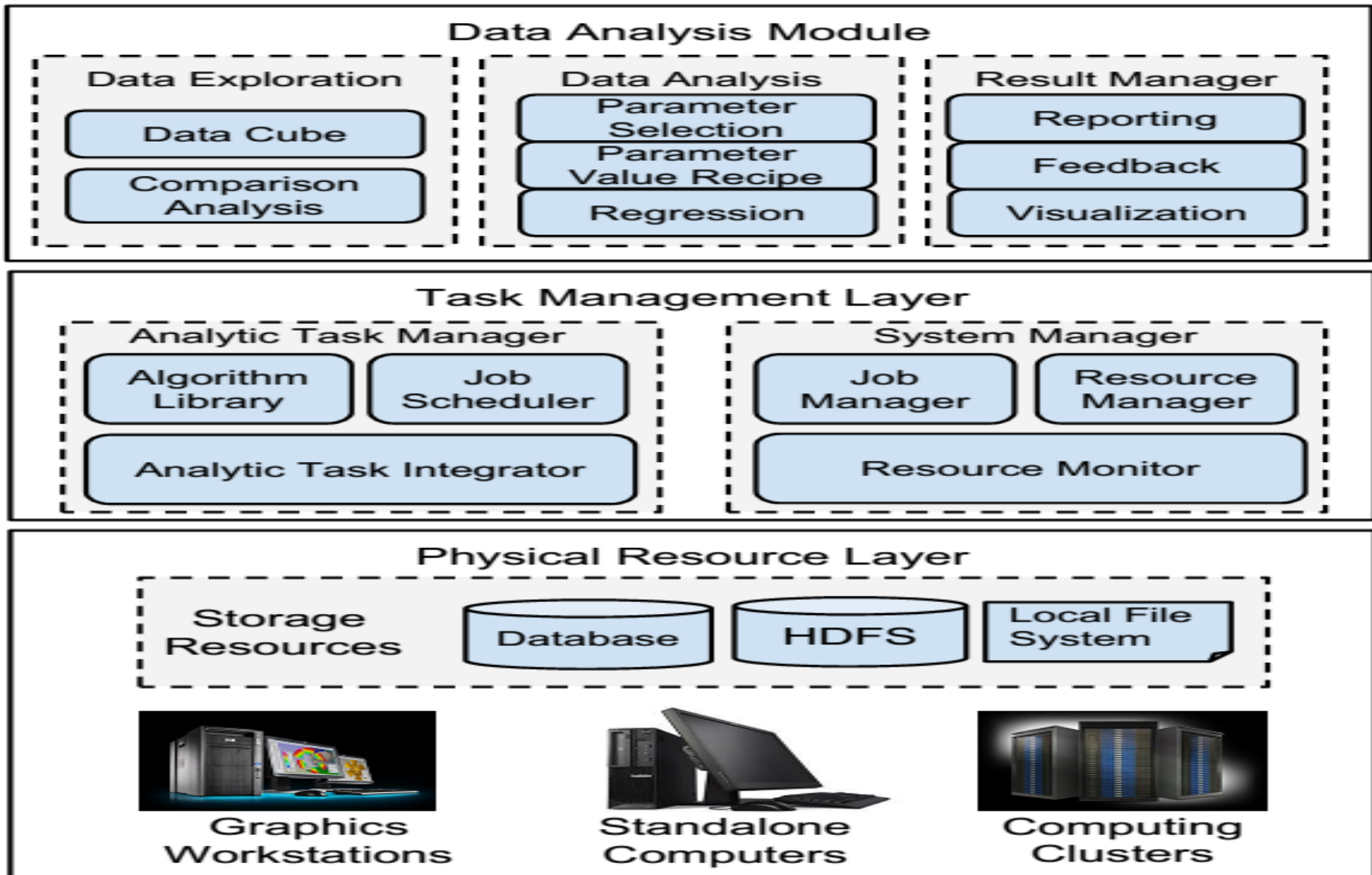
- **Solution:**
 - ✓ Come up with an ensemble feature selection algorithm;
 - ✓ Utilize the regression model to quantify the correlation between product quality and various parameters.
 - ✓ Apply association based method to identify feature combinations which can improve the quality of product.

Methodologies

PDP-Miner, is designed, implemented and deployed as an integrated data analytics platform customized for process optimization in PDP manufacturing.

- ❑ It provides a more convenient way to facilitate manufacturers to conduct PDP production data analysis.
- ❑ Based on a distributed system, it supports high-performance analysis with large scale production data.
- ❑ It extends and develops appropriate data mining algorithms and adapts them to the problem of analyzing the manufacturing data.

System Overview



Data Exploration

Data Exploration



(a) Comparison Analysis



(b) Data Cube

Result Management

Result Management

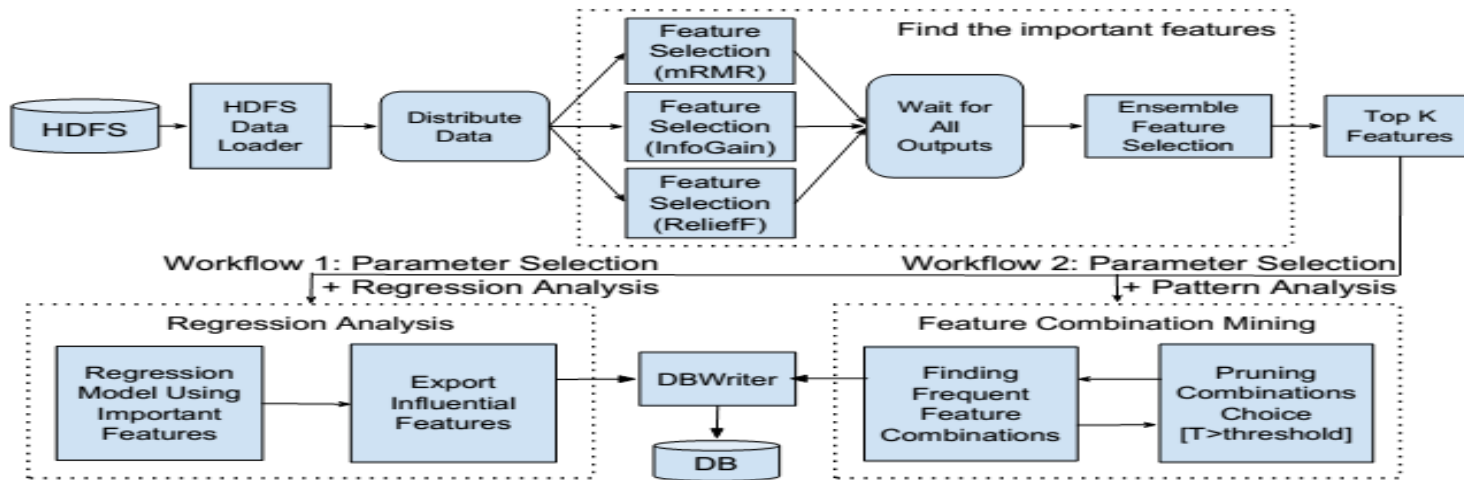


(g) Visualization



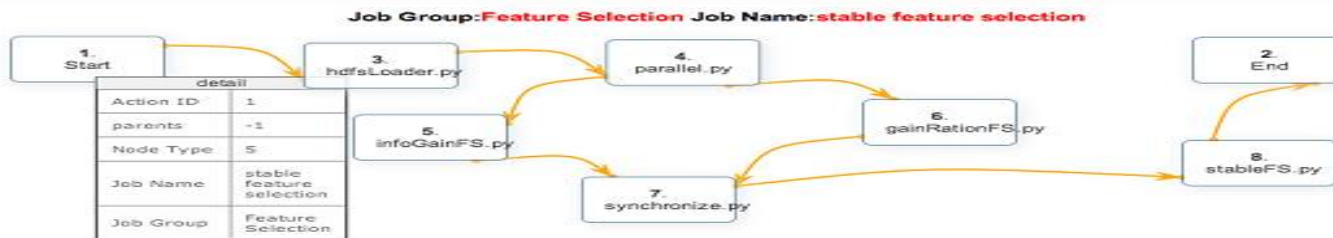
(h) Result List , Feedback Collector

Data Analysis



Schedule	Actions	Group	Name	Description	Status	Latest Update
		test	1	A simple sequential order configuration	unscheduled	2014-02-20 14:22:40
		Classification	CAR		unscheduled	2013-11-13 16:31:04
		Clustering	KMeans Algorithm	KMeans algorithm for clustering	unscheduled	2013-11-13 17:22:07
		Feature Selection	stable feature selection	based on multiple feature selection algorithms, stable features are extracted.	scheduled	2014-02-12 17:44:57

Showing 1 to 4 of 4 records



Data Analysis

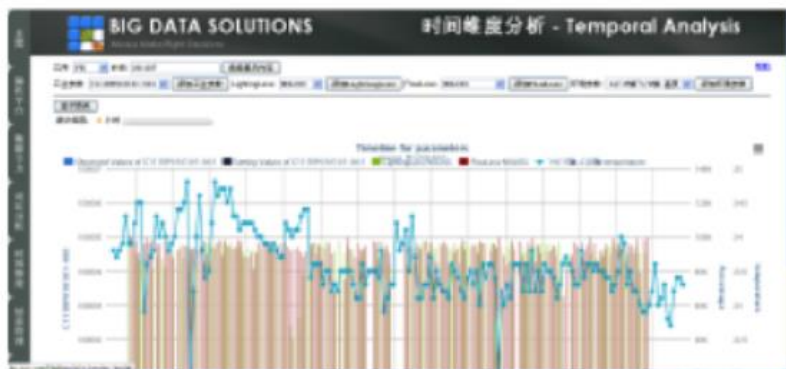
Data Analysis



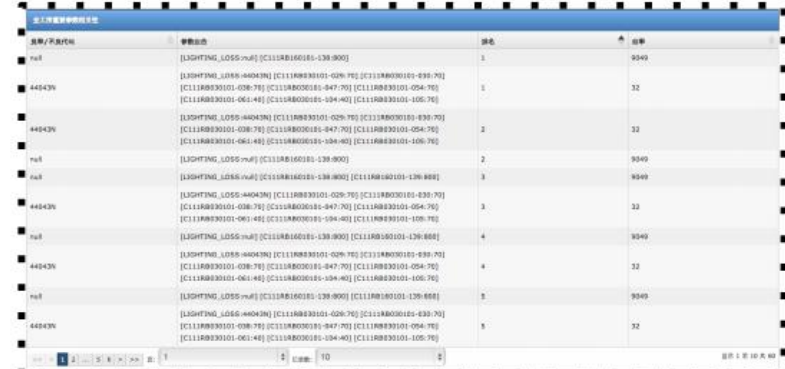
(c) Operation Panel



(d) Parameter Selection



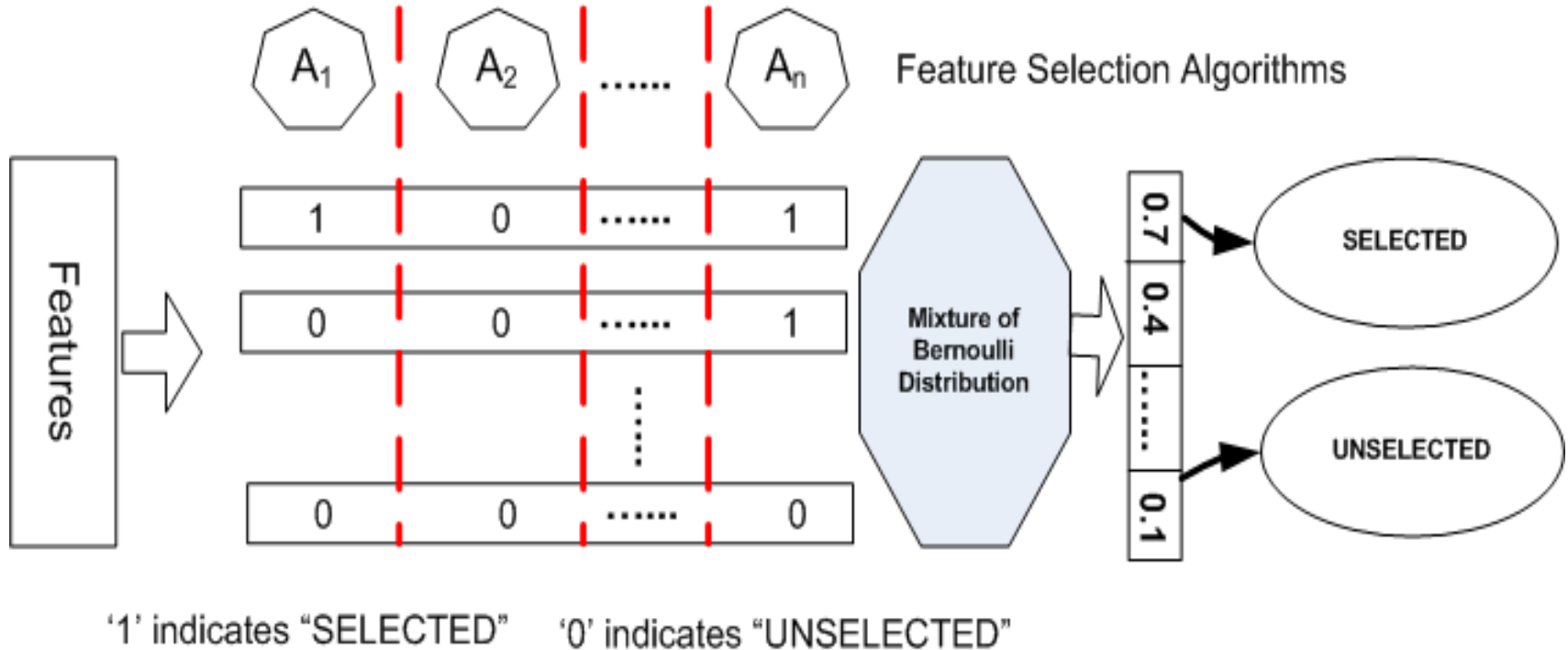
(f) Regression Analysis



(e) Discriminative Analysis

Ensemble Feature Selection

- Each Feature can be selected or unselected.
- Different feature selection algorithms get different results.
- Stable feature selection is achieved by EM.



Regression Analysis

- Linear regression model is used to establish the relationship between important controlling parameters and production rate.
 - Each parameters value are been normalized.
 - The weights of parameters in the model can be easily interpreted.
 - $|w_i|$ shows the conditional correlation between feature i and production rate.
 - The p-value show its likelihood of the correlation.

Discriminative Analysis

- Aim: identify some combinations of features which are closely related to the qualified panels or defective panels.
- Association based classification is used. The rule set is selected if the feature set can differentiate qualified and defective panels.

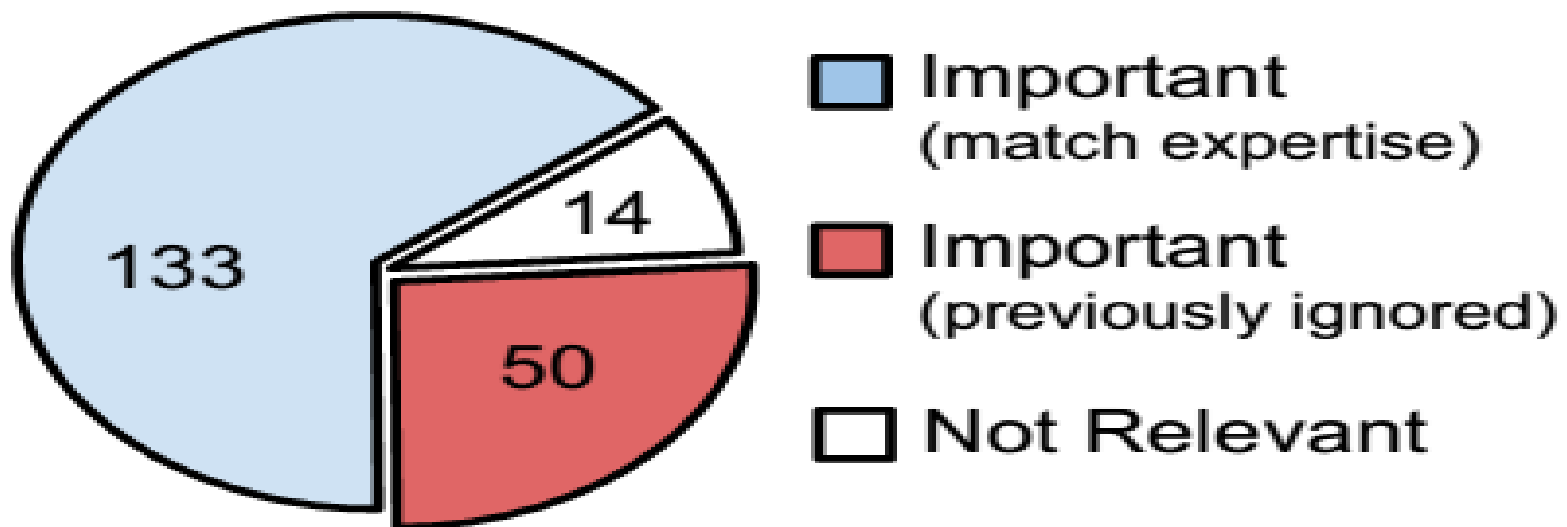
$$DisS(\alpha) = |S_{qualified}(\alpha) - S_{defective}(\alpha)|$$

System Deployment

- We deploy our system for ChangHong Company, one of the world's largest display device manufacturing companies in China.

Real Findings(by feature selection)

- 197 important parameters are reported by our system.
- 133 important parameters are consistent with the production experience
- 50 important parameters have been verified by domain experts to have direct impact on production quality.
- 14 parameters are not relevant.

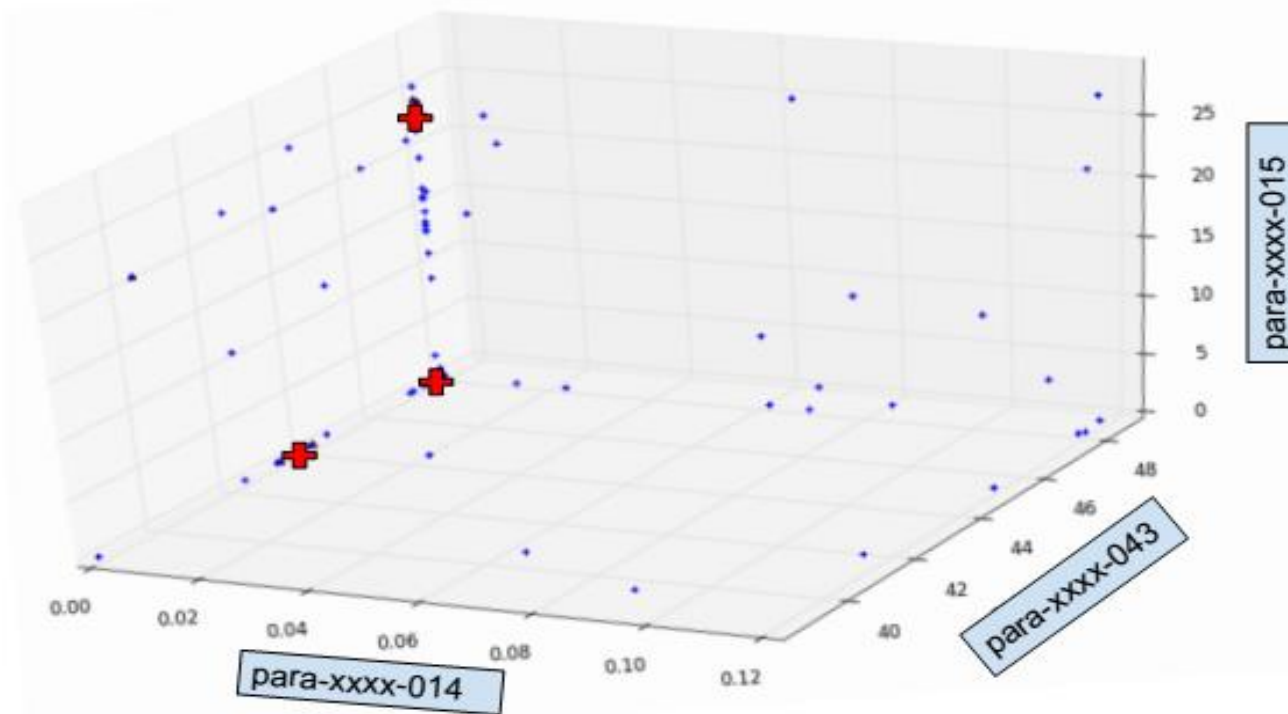


Real Findings(by regression)

- The variance of humidity plays an important role in affecting the yield rate. It has negative correlation with yield rate.
- The less the pressure changes, the higher the yield rate would be.
- The temperature is controlled very well under 27°C. Its weight in regression model is small.

Real Findings(by Discriminative Analysis)

- $\langle \text{para-xxxx-014} = 0, \text{para-xxxx-015} = 0 \text{ or } 24, \text{para-xxxx-043} = 44 \text{ or } 48 \rangle$ have large support for defective panel. Such parameter value combination should be avoided.



Conclusion

- By taking advantage of our system, the overall PDP yield rate increases from 91% to 94%. Monthly production capacity is boosted by 10,000 panels, which brings more than 117 million RMB of revenue improvement per year.

Question and Answer

- Thanks