転移学習に基づくケミカルトナー製造工程における品質予測

Transfer Learning for Quality Prediction in a Chemical Toner Manufacturing Process

小林　翔太* Shota KOBAYASHI
宮川　将士* Masashi MIYAKAWA
武政　奨* Susumu TAKEMASA
高橋　直希* Naoki TAKAHASHI
渡邉　好夫* Yoshio WATANABE
佐藤　敏明** Toshiaki SATOH
加納　学*** Manabu KANO

要　旨

ケミカルトナー製造工場では，トナー品質や生産性の向上のために，設備や原材料の変更が頻繁に行われる．これらの変更に伴い，トナーの自動品質制御システムの重要な役割を担う予測モデルの再構築が必要となり，システムダウンの原因となる．このダウンタイムを短縮するために，設備や原材料の変更直後に得られる少数のデータから高精度な予測モデルを構築できる，転移学習を活用した効率的なモデリング手法を開発した．これは，Frustratingly Easy Domain Adaptationを拡張した，新たな異種ドメイン適応手法である．予測には，モデルの汎用性と精度を向上させるため，バギングを用いたガウス過程回帰（GPR）を採用した．提案手法は，部分最小二乗回帰，ランダムフォレスト，GPRと比較して，優れた性能を示した．この手法をトナー量産工場に適用した結果，全てのトナー品質で予測精度目標を満足することができ，トナー品質管理者の工場管理工数の75%削減を達成した．

ABSTRACT

In chemical toner manufacturing plants, equipment and raw materials are frequently changed to improve the toner quality and productivity. These changes require reconstruction of the prediction model, which plays a key role in the automatic quality control system, and cause downtime. To reduce the downtime, we developed an efficient modelling method based on transfer learning, which can build an accurate model from small-size data obtained just after the changes. By extending Frustratingly Easy Domain Adaptation, a new heterogeneous domain adaptation technique was proposed. In addition, gaussian process regression (GPR) was adopted with bagging to improve the robustness and accuracy of the model. The proposed method showed superior performance to partial least squares regression, random forest, and GPR. Finally, the proposed prediction method was applied to a toner mass-production plant; the prediction accuracy target was satisfied for all toner qualities. As a result, a 75% reduction in plant control person-hours of the toner quality manager was achieved.

*　CMC事業本部　第二トナー事業センター
2nd Toner Business Center, CMC Business Division
**　株式会社　明和eテック
MEIWA e-TEC Company, LTD.
***　京都大学大学院　情報学研究科　システム科学専攻
Department of Systems Science, Kyoto University

Reprinted from Computer Aided Chemical Engineering 49, Kobayashi, S., Miyakawa, M., Takemasa, S., Takahashi, N., Watanabe, Y., Satoh, and T., Kano. M., “Transfer Learning for Quality Prediction in a Chemical Toner Manufacturing Process”, 1663-1668 (2022), with permission from Elsevier.

1. Introduction

In recent years, automatic quality control has been used for stabilizing chemical toner quality and determining efficient operating conditions in toner plants[1,2]. In these plants, equipment and raw materials are often changed. As shown in Fig. 1, such a change alters the dimensions and distributions of input variables, makes it difficult to use existing prediction models, and makes it necessary to reconstruct the models. During the re-accumulation of training data, automatic quality control is forced to stop functioning, and manual quality control is required. This manual control requires many person-hours, therefore, it has been desired to construct an accurate prediction model using as short-term data as possible.

A promising approach to solve this problem is to use transfer learning. We expanded Frustratingly Easy Domain Adaptation (FEDA), which is a simple homogeneous domain adaptation method, to cope with a heterogeneous domain adaptation (HDA) problem without complex parameter tuning. The proposed method is referred to as Frustratingly Easy Heterogeneous Domain Adaptation (FEHDA). Moreover, we utilized a combination of Gaussian Process Regression (GPR) and bagging, a type of ensemble learning, for predicting the toner quality.

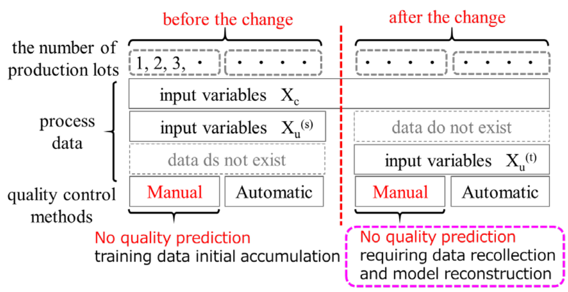

Fig. 1 The influence of changes in equipment and raw materials.

2. Chemical toner manufacturing process

The chemical toner manufacturing process treats one lot per day, and it takes seven days from raw material to final product, as shown in Fig. 2. The IoT-based manufacturing process data collection system handles several thousand variables (items), including raw material properties, equipment operation conditions, and toner quality, and stores data of several hundred lots or more.

Before the introduction of automatic quality control, toner quality was controlled manually by the toner quality manager, who determined the optimum operating condition for lot based on the quality measurements of the lots whose manufacturing *N* was finished (lot *N*-2 and older). The manual quality control consumes many person-hours and increases the risk of out-of-specification due to variations in toner quality.

The automatic quality control system currently in operation consists of a quality prediction module that predicts future toner quality and an operating condition optimization module that determines the operation amount[2].

This automatic quality control is feed-forward inferential control, which simulates the manual operation, as shown in Fig. 2. However, as mentioned in the previous section, the change of equipment and raw materials requires the data re-accumulation of at least 40 lots (days) so that the model reconstructed satisfies the accuracy target.

3. Prediction using transfer learning

3-1 Frustratingly Easy Heterogeneous Domain Adaptation

FEDA is a method of transfer learning that is easy to implement with simple feature space expansion[3].

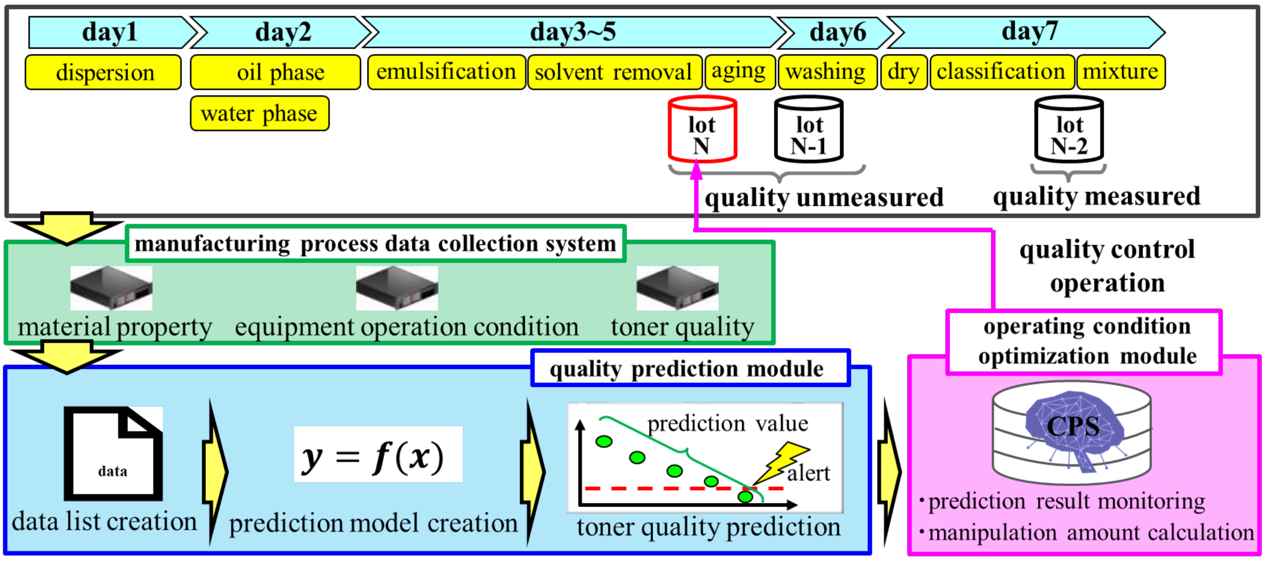

Fig. 2 Chemical toner manufacturing process and automatic quality control system[2].

Assuming that the input variables $\boldsymbol{x}^{(s)}$ in the source domain (hereinafter referred to as "SD") and $\boldsymbol{x}^{(t)}$ in the target domain (hereinafter referred to as "TD") are K-dimensional, the input variables in both domains are expanded into $3K$-dimensional features as follows:

$$D_s = \left(\boldsymbol{x}^{(s)},\ \boldsymbol{x}^{(s)},\ 0\ \right) \tag{1}$$
$$D_t = \left(\boldsymbol{x}^{(t)},\ 0\ ,\ \boldsymbol{x}^{(t)}\right) \tag{2}$$

The expanded feature space consists of a space with features common to both domains, a space with features unique to SD, and a space with features unique to TD. Also, $0 = (0, 0, 0, \ldots, 0) \in \mathcal{R}^K$ in Eqs. (1) and (2) is the zero vector.

In the manufacturing process, due to changes in equipment and raw materials, the configuration of the manufacturing equipment differs in both domains, which makes the location and number of installed sensors also different. Hence, when heterogeneous domain adaptation is required, FEDA cannot be used as it is. To make FEDA applicable to heterogeneous domain adaptation (HDA), heterogeneous feature augmentation (HFA) was proposed[4]. This method needs much computational time because to solve an optimization problem for finding the optimal latent space.

We propose frustratingly easy heterogeneous domain adaption (FEHDA), which is a direct and simple extension of FEDA and applicable to HDA. The proposed method does not require solving the optimization problem. We divide input variables $\boldsymbol{x}^{(s)} \in \mathcal{R}^P$ in SD into ${\boldsymbol{x}_c}^{(s)} \in \mathcal{R}^K$ that is common to SD and TD and ${\boldsymbol{x}_u}^{(s)} \in \mathcal{R}^{P-K}$ that is unique to SD. Similarly, input variables $\boldsymbol{x}^{(t)} \in \mathcal{R}^Q$ in TD is divided into the common input variables ${\boldsymbol{x}_c}^{(t)} \in \mathcal{R}^K$ and the unique input variables ${\boldsymbol{x}_u}^{(t)} \in \mathcal{R}^{Q-K}$. As shown in Fig. 3, ${\boldsymbol{x}_c}^{(s)}$ and ${\boldsymbol{x}_c}^{(t)}$ are expanded as in Eqs. (1) and (2), respectively, while ${\boldsymbol{x}_u}^{(s)}$ and ${\boldsymbol{x}_u}^{(t)}$ are placed in the space with unique features in each domain as follows:

$$D_s = \left({\boldsymbol{x}_c}^{(s)},\ {\boldsymbol{x}_c}^{(s)},\ {\boldsymbol{x}_u}^{(s)},\ \ 0\ \ ,\ \ 0\ \right) \tag{3}$$
$$D_t = \left({\boldsymbol{x}_c}^{(t)},\ \ 0\ \ ,\ \ 0\ \ ,\ {\boldsymbol{x}_c}^{(t)},\ {\boldsymbol{x}_u}^{(t)}\right) \tag{4}$$

3-2 Prediction Model

To build a prediction model, we propose a method that combines Gaussian process regression (GPR) and bagging. The input variables are the expanded ones in Eqs. (3) and (4). GPR can predict not only the expected values but also the standard deviations of output variables and provide the reliability of the prediction. Bagging is a form of ensemble learning that uses bootstrap sampling to

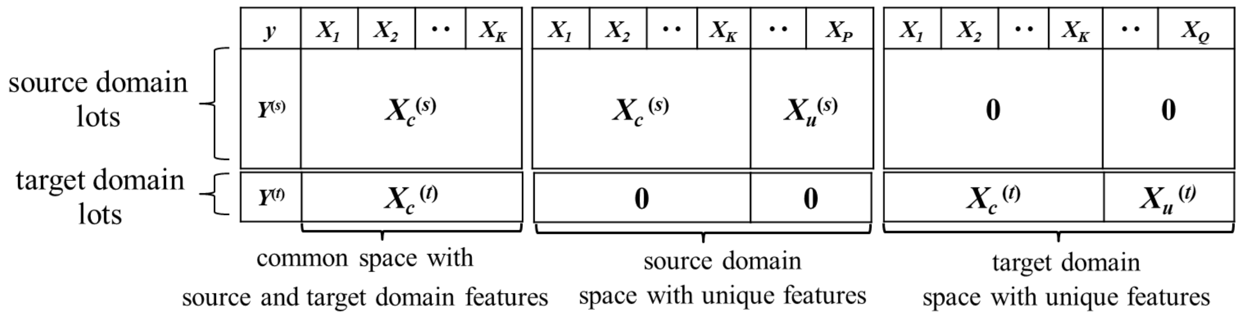

Fig. 3 Feature space expansion in Frustratingly Easy Heterogeneous Domain Adaptation (FEHDA).

construct many independent weak learners and then integrates the results of the weak learners into a prediction. Kamishima et al. (2009) proposed TrBagg[5)], which uses bagging for transfer learning. TrBagg builds weak learners using data sampled from SD and TD. The weak learners are adopted based on the classification errors for TD. The method may cause over-fitting or require the separation of TD for validation.

The chemical toner manufacturing process produces only one lot per day. To reduce the downtime of the automatic quality control system, the number of samples after each change, which are used for reconstructing the prediction model, needs to be limited. That means the number of TD lots must be small, i.e., about 10 lots. Since TrBagg does not work well in such a situation, we did not adopt it. In the proposed method, bagging is modified by selecting only weak learners with small standard deviations of the output variables when integrating the results of the weak learners. The weak learners with small standard deviations are expected to give a more reliable prediction because it is considered to use data with high similarity to the target lot preferentially. We use sequential updating of the prediction model for each lot.

4. Comparison of prediction methods

The proposed modeling method, i.e., GPR and bagging, was compared with the typical regression methods, partial least squares regression (PLSR), random forest (RF), and GPR in two cases: 1) change of coloring materials, representing material improvement, and 2) change of production scale, representing equipment improvement. FEHDA was used in both cases, and the two most important qualities were investigated. The dimensions of the input variables are shown in Table 1.

In case 1, black and magenta toners, which were made from almost the same materials except for the coloring one, were targeted, regarding black toner as SD and magenta toner as TD. In case 2, the same color toner manufactured by equipment with different scales was targeted, regarding the large scale plant as SD and the small scale one as TD.

Table 1 The dimensions of the input variables and the number of lots in two cases: 1) change of coloring materials and 2) change of production scale.

	input variables	case (1)	case (2)	Application to a mass production plant
the number of variables	$X^{(s)}$	2821	2826	2914
	$X^{(t)}$	2831	2266	2721
	$X_c^{(s)}$, $X_c^{(t)}$	2789	1803	2598
	$X_u^{(s)}$	32	1023	316
	$X_u^{(t)}$	42	463	123
the number of lots	$X^{(s)}$, $X_c^{(s)}$, $X_u^{(s)}$	450	450	333
	$X^{(t)}$, $X_c^{(t)}$, $X_u^{(t)}$	10-109	10-109	10-39

The prediction accuracy was evaluated using Root Mean Squared Error (RMSE). In defining the target for the prediction accuracy, the following conditions were set; first, the center of the predicted distribution of the qualities is within 50% of the process specification width Δ, and second, the probability of out-of-specification is less than 0.3% when the quality prediction value is at the upper or lower limit of the process specification width Δ. The variability was assumed to be normally distributed[2]. Based on these conditions, the target value for prediction accuracy became 0.5Δ≥6 RMSE, i.e., RMSE/Δ≤8.3[%].

Fig. 4 shows the evaluation results for the 11th to 110th lots in TD. The proposed method outperformed the other methods in both cases and also satisfied the prediction accuracy target. In particular, a more significant improvement was obtained in case 1. Fig. 5 shows the predicted and measured values for each lot of quality #1. It was confirmed that the predicted values followed the trend of the actual measured values, and there were no large errors in all lots. On the other hand, the improvement achieved by the proposed method in case 2 was smaller than that in case 1. This can be attributed to the large proportion of intrinsic variables that account for 45% of the input variables in each domain, which implies that SD contains less valid information for the transfer.

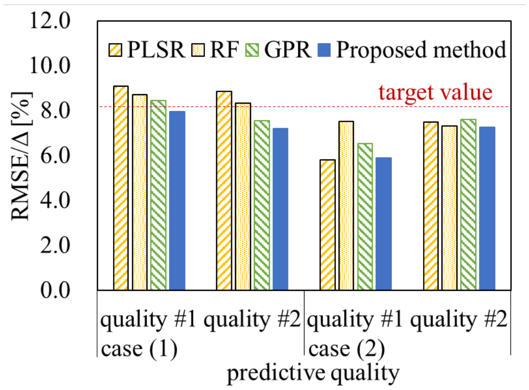

Fig. 4 Comparison of RMSE's of prediction methods.

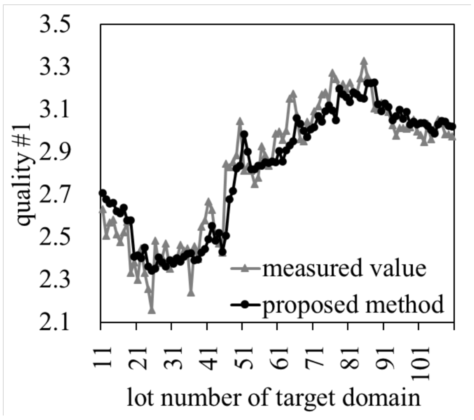

Fig. 5 Quality #1 prediction results of the proposed method for case 1.

5. Application to a mass-production plant

The proposed method was applied to a mass-production plant in RICOH. There are 12 quality items to be predicted, including particle size distribution, particle shape, and charging characteristics. The numbers of variables and lots are shown in Table 1. The proposed prediction method was compared with two different methods using only TD (hereinafter, referred to as Target) and using only common input variables in SD and TD (referred to as Common). In these two methods, we used random forest, which has been used in the existing automatic quality control[2].

We conducted the prediction of the 12 qualities from the 11th lot to the 40th lot in TD. While Target and Common failed to achieve the prediction accuracy target for two and three quality items, respectively, the proposed method achieved the prediction accuracy target for all quality items. Besides, the proposed method outperformed Target and Common in all qualities. The prediction accuracy in RMSE of the proposed method was 11.4% higher than Target on average, and particularly 17.4% for quality #10. Compared to Common, the average improvement was 15.4%, and the best was 25.4% in quality #4.

Fig. 6 shows the predicted and measured values for each lot of quality #2. The predicted values of the proposed method follow the measured values better than those of Target and Common. In particular, the prediction errors in the initial stage for lots 18, 24, 25, and 26 are small. The prediction accuracy indices, i.e., RMSE/Δ, are 8.1% for the proposed method, 9.5% for Target, and 8.6% for Common, indicating that the proposed method satisfies the prediction target values for these small lots.

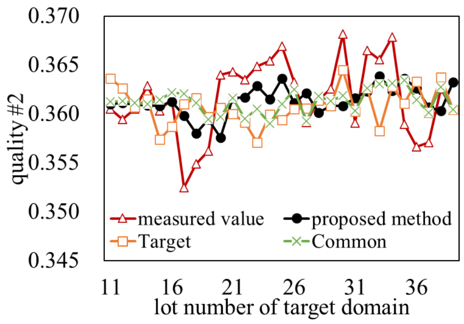

Fig. 6 The prediction results of the proposed method, Target, and Common for quality #2.

The toner qualities predicted by the transfer learning were used in the automatic quality control system based on feed-forward inference control[2)] described in Section 2. Before applying prediction by the transfer learning, 40 lots (days) of data had to be accumulated to achieve the required accuracy target. With the proposed method, the data accumulation was reduced to 10 lots (days), and the person-hours required for monitoring and control by quality managers immediately after a change in equipment or raw materials were reduced by 75%.

6. Conclusions and future tasks

We first proposed a new transfer learning method that can cope with heterogeneous domain adaptation, i.e., FEHDA, which is simple extension of FEDA. Second, we proposed a new prediction method that combines Gaussian process regression (GPR) and bagging. Finally, the proposed method was adopted in the automatic control system of RICOH's chemical toner plant. The downtime of the automatic quality control system decreased from 40 lots (days) to 10 lots (days), and the person-hours required for manual quality control by toner quality managers have been reduced by 75%.

References

1) H. Khorami et al.: Multivariate Modeling of a Chemical Toner Manufacturing Process, *Chemical Engineering and Technology*, Vol. 40, pp. 459-469 (2017).

2) N. Takahashi et al.: Automatic Quality Control for Chemical Toner Using Machine Learning Prediction Technology, *Society of Instrument and Control Engineers 7th Multi Symposium on Control Systems 2A2-3*, in Japanese (2020).

3) H. Daumé III: Frustratingly Easy Domain Adaptation, *Proceedings of the 45th Annual Meeting of ACL*, pp. 256-263 (2007).

4) L. Duan, D. Xu, I.W. Tsang: Learning with Augmented Features for Heterogeneous Domain Adaptation, *Proceedings of the 29th International Conference on Machine Learning*, ICML 2012 1, pp. 711-718 (2012).

5) T. Kamishima, M. Hamasaki, S. Akaho: TrBagg: A simple transfer learning method and its application to personalization in collaborative tagging, *in: Proceedings - IEEE International Conference on Data Mining*, ICDM, pp. 219-228 (2009).