

機械学習を用いた工程情報に基づく作業時間のバラツキ要因の特定

山城 広周[†] 野中 尋史[†] 井若 玄貴[†]

[†]長岡技術科学大学 〒940-2137 新潟県長岡市上富岡町 1603-1

E-mail: [†] {s173358, nonaka, s161006}@{stn, kjs, stn}.nagaokaut.ac.jp

あらまし 製造業において、作業時間分析を行うことは生産性の向上の観点で極めて重要となる。現在、作業時間を自動的に分析する手法として画像と機械学習を用いた研究が行われている。一方、材料種別や作業 ID などの工程情報は画像にはない情報を含み、詳細な作業分析を行うことができる利点がある。そこで、工程情報のカテゴリデータ、及び数値データを用いて作業時間のバラツキの要因推定を行う手法を提案する。機械学習モデルから予測された作業時間と実作業時間の誤差の大きい作業をバラツキの大きい作業と定義した。そのうえで、作業時間のバラツキ要因推定には Kullback-Leibler (KL) divergence を用いた手法を提案する。また、本研究の提案手法の検証として、実際の製造工場の工程情報を用いて検証を行った。

キーワード 機械学習, 作業工程分析

1. はじめに

製造業の現場において、作業時間の効率化を目的とした作業分析は極めて重要なタスクとなる[1, 2]。しかしながら、現状は手動でカメラなどのデータを解析する手法がメインであり、多大な労力・コストを費やす必要がある。そこで、作業分析を自動的に行う手法の開発が求められている。そのよう中、作業を自動的に分析する手法としてウェアラブルデバイスによるセンシングや画像と機械学習を用いた研究が行われている[3, 4, 5]。例えば、兵頭ら[3]はリストバンド型のウェアラブルデバイスからセンシングされたデータや生産管理システムのデータを分析することで、作業内容や作業員の作業処理時間に関して分析を行った。画像を用いた例として、高橋ら[5]は作業員の姿勢からその作業が作業に含まれる無駄かどうかを分類するモデルを提案している。一方で材料の種別や作業 ID などの工程情報はセンサーデータや画像にない情報も含み、詳細な作業分析を行うことができる利点がある。しかしながら、工程情報に基づき作業時間の無駄やバラツキの要因を分析する研究は筆者が知る限り行われていない。

そこで、本研究では工程情報に含まれるカテゴリデータ、及び数値データを用いて作業時間のバラツキの要因分析を行う。具体的には作業時間のバラツキが大きい群に影響を与える要因について、Kullback-Leibler (KL) divergence を用いてバラツキが小さい群と各項目との距離を計算する手法を提案する。提案手法の検証として、企業より頂いた実際の作業工程のデータを用いて要因分析を行う。

2. 手法

本研究では、作業時間のバラツキの要因分析を行うための手法として、Kullback-Leibler (KL) divergence (KL 情報量) を用いる。

実作業時間と何かしらの手法によって予測した予測値の差を誤差として算出する。次に、誤差が平均以下かどうかで True, False のフラグを立て、各項目の各要素も One-hot ベクトルに変換し、誤差が平均以下のベクトルと各要素のベクトルで Kullback-Leibler divergence を計算して分析に用いる。本研究で用いる KL 情報量は以下の式で算出する。

$$D_{KL}(p||q) = \sum_i^N p(x_i) * \log \left(\frac{p(x_i)}{q(x_i)} \right) \quad (1)$$

KL 情報量を用いた手法に関して、図 1 に概略図を示す。この手法を用いることで、任意の要素が誤差の小さい集合に偏って出現するか計算することが可能である。

誤差平均: 220

誤差	200	100	300	100	150	400	200	240	300	200
平均以下 (One-hot)	1	1	0	1	1	0	1	0	0	1
要素A	1	0	0	0	0	0	1	1	0	0
確率化										
平均以下	0.6	0.6	0.4	0.6	0.6	0.4	0.6	0.4	0.4	0.6
要素A	0.3	0.7	0.7	0.7	0.7	0.7	0.3	0.3	0.7	0.7

Kullback-Leibler divergence

0.29634501697786236

図 1. ハミング距離の概略図

3. 実験

本研究では、提案手法を実際に用いるため、研究協力企業より提供された製造サンプルごとにまとめられた工程情報を使用する。本研究で用いた工程情報を表 1 にまとめる。また、各工程の項目の要素数を表 2, 3 にまとめる。

表 1. 製造サンプルの項目数とデータ数

工程	項目数	データ数
A	12	36139
B	20	22236

表 2. 工程 A の項目ごとの要素数

項目	要素数
作業種	2
機械種	22
作業者	16
作業完了	2
出来映え	5
材料 A	155
材料パラメータ A	55
材料パラメータ B	94
材料パラメータ C	2
材料パラメータ D	18
材料パラメータ E	22

表 3. 工程 B の項目ごとの要素数

項目	要素数
作業種	3
機械種 A	13
機械種 B	8
作業者	18
作業完了フラグ	2
出来映え	6
材料 A	113
材料パラメータ A	47
材料パラメータ B	87
材料パラメータ C	2
材料パラメータ D	20

材料パラメータ E	5
材料パラメータ F	16
材料パラメータ G	21

図 2 は実験の概要図である。手法は大きく 2 つの手順に分かれる。初めに、機械学習モデルを用いて実作業時間を予測し、実作業時間と予測した作業時間を用いて作業時間の誤差を計算する。次に、KL 情報量を用いて作業時間のバラツキが大きく、機械学習モデルにも影響を与えている工程情報に関する要因を抽出する。



図 2. 実験の概要図

3.1. 作業時間の予測

作業時間の予測には、機械学習の Boosting Tree 系のモデルの 1 つである LightGBM[4] を用いる。図 3 は LightGBM のモデル概要図である。LightGBM では、入力データのヒストグラムからサンプリングを行い、そのデータに重みを付けて単純な決定木に入力とする。各決定木の学習はシーケンシャルに行い、入力データの重みは前回までの学習結果を反映させる。LightGBM は、一般的な機械学習モデルと比較して訓練時間が短く、推測精度が高いため、本研究で作業時間予測手法として採用した。また、本研究では用いた項目のうち、作業時間以外を入力として、作業時間を予測した。

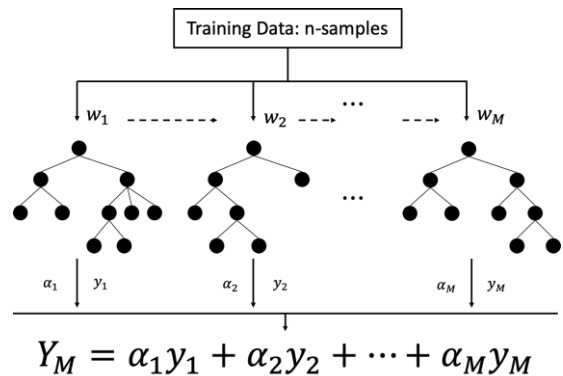


図 3. LightGBM のモデル概要図

4. 結果

表 2 に LightGBM の推論精度を示す。予測モデルの評価として、二乗平均平方根誤差(RMSE)を用いて評価

を行った。

表 4. LightGBM の RMSE

工程	RMSE	実測値の平均
A	189.80	1327.35
B	319.67	1609.86

また、表 3, 4 にバラツキが大きい群に頻出する項目 (KL 情報量の上位 5 つ) の計算結果を示す。要素名の “other” は、出現回数が 10 回以下の要素をまとめたものである。

表 5. 工程 A の KL 情報量上位 5

項目	要素名	KL 情報量
材料パラメータ 1	15	13673.11
材料パラメータ 1	98	13646.56
材料パラメータ 2	Other	13637.84
出来映え	Other	13630.20
材料パラメータ 1	108	13614.70

表 6. 工程 B の KL 情報量上位 5

項目	項目名	KL 情報量
機種	Other	7873.68
名前 CD	76221	7873.68
材料パラメータ 1	15	7873.68
材料パラメータ 2	Other	6856.56
計量器 CD	Other	7818.96

5. おわりに

本研究では、製造工場の工程情報を用いた作業時間のバラツキの要因分析を行う手法を提案した。具体的な手法として、Kullback-Leibler (KL) divergence を用いた。実作業時間と予測作業時間の差を計算し、誤差が平均以下の群と各項目の間で Kullback-Leibler (KL) divergence を計算することで、標準的な作業時間からのバラツキが小さい項目は値が小さく、バラツキが大

きい項目は値が大きくなることでどの項目がバラツキの要因となるのか計算可能となる。

本研究では、提案手法の検証として、研究協力企業の製造工場の工程情報に適用させた。本研究では、2 つの工程に関して、それぞれ 36139 件、12 項目、393 要素と 22236 件、20 項目、369 要素のデータに適用した。

工程情報のデータに対し、作業時間を予測するモデルとして LightGBM を用いて作業時間の推定を行い、そこから計算された誤差をもとに Kullback-Leibler (KL) divergence を計算した。

今後の課題として、抽出した作業時間のバラツキ要因に関して提供元企業と評価を行うことがあげられる。

参考文献

- [1] Lavoie, P., Kenné, J. P., & Gharbi, A. (2009). Optimization of production control policies in failure-prone homogenous transfer lines. *IIE Transactions*, 41(3), 209-222.
- [2] Cus, F., & Balic, J. (2003). Optimization of cutting process by GA approach. *Robotics and Computer-Integrated Manufacturing*, 19(1-2), 113-121.
- [3] 兵頭靖得, 小橋武弘, & 山中泰介. (2016). IoT 行動センシングを用いた作業分析技術. *東芝レビュー*, 71(5), 72-75.
- [4] 福永いづみ, & 井本和範. (2018). 物流現場における生産性向上システムのための作業分析. In 年次大会 2018 (p. J1630107). 一般社団法人 日本機械学会.
- [5] 高橋典宏, 山澤一誠, 生雲公啓, 野田賢, & 横矢直和. (2007). 距離画像センサを用いた俯瞰画像からの SVM による人物の姿勢分類. *電子情報通信学会技術研究報告*. 電子情報通信学会, 107(384), 47-52.
- [6] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, 3146-3154.