

機械学習とエントロピーを用いた作業時間の無駄に関する要因分析

長岡技術科学大学 *山城 広周 YAMASHIRO Hirochika
 長岡技術科学大学 井若 玄貴 IWAKA Genki
 長岡技術科学大学 野中 尋史 NONAKA Hirofumi

1. はじめに

製造業の現場において、数理最適化手法の適用が従来から行われてきた。最適化が行われてきた例として、生産ラインの最適化[1]、製造機械のパラメータの最適化や製造コストの最適化[2]などがある。そのような中、画像を利用した作業最適化手法の研究が提案されている。例えば、高橋ら[3]は画像を用いて作業者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。一方で材料種別や作業者 ID などの工程情報は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、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を回帰モデルとして高いパフォーマンスを示す LightGBM を用いて工程情報から作業時間を予測した上で、作業時間のバラツキを LightGBM で予測した作業時間（標準的な作業時間）と実測値との差と仮定する。仮定に基づき、標準作業時間である予測値と、実測値との差（誤差）を計算し、差が大きい、すなわち、バラツキが大きい群に影響を与える要因についてエントロピーを用いて抽出する手法を提案する。

2. 手法

本項では、提案手法に関して具体的に述べる。図1は本研究での手法の概要図である。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、実作業時間と予測時間を用いて作業時間の誤差を計算する。次に、エントロピーを用いてバラツキが大きいデータに関する要因を抽出する。

2.1 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2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。



図1. 手法の概要図

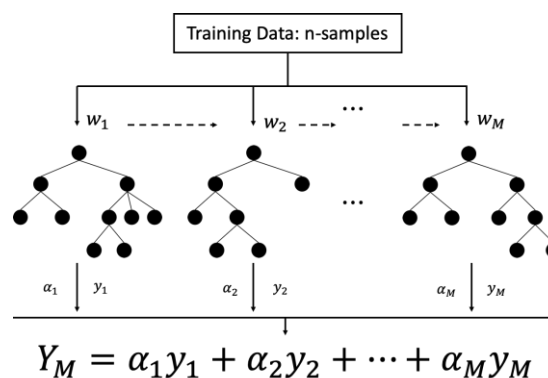


図2. LightGBM のモデル概要図

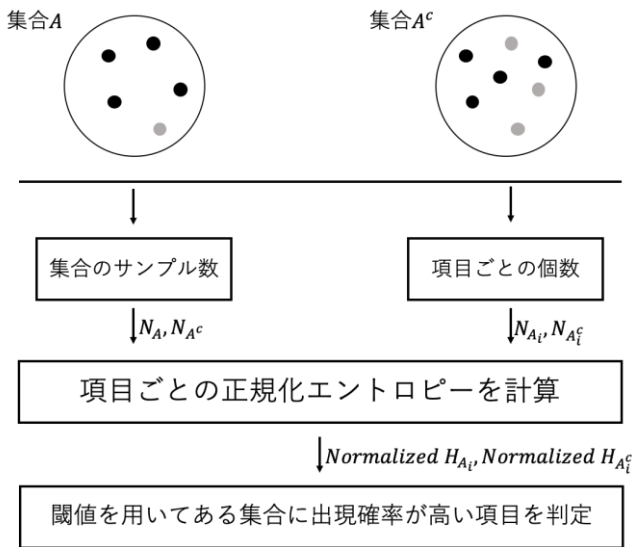
2.2 要因分析

製造個数単位の実作業時間と前項で計算した予測値の差を誤差として算出する。次に、誤差の大きいデータ群、小さいデータ群にデータを分け、誤差の大きいデータ群によく出現する項目を算出する。項目の算出にはエントロピーを用いて誤差の大きいデータ群、小さいデータ群のエントロピーの差が2倍以上あるかを閾値として判定する。エントロピーを用いた手法に関して、図3に概略図を示す。この手法を用いることで、任意の項目がある集合に偏って出現するか計算することが可能で、かつ、正規化を行うことでデータ数に依らない比較を行うことができる。

エントロピーに関しては以下の式を用いて行う。

$$Normalized H_{A_i} = -\frac{\left(\frac{1}{N_A} \log_2 \frac{1}{N_{A_i}}\right)}{\log_2 N_A} N_{A_i} \quad (1)$$

N_A は、バラツキの大きい、もしくはバラツキの小さい集合のデータ数であり、 N_{A_i} は、集合内で任意の*i*番目の項目を含むデータが何件存在するかを表している。



3. 実験

本研究では、A社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。本研究では2つの工程に関する、データを利用した。製造サンプルは工程Aが31137件、工程Bが15270件である。また、工程項目は作業者ID、商品ID、材料、材料パラメータ、作業時間、製造個数、製品質量などの16件である。作業時間以外を入力として、作業時間を予測した。

4. 結果

表1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 1. LightGBM の RMSE

工程	RMSE	実測値の平均
A	179.25	1256.42
B	65.07	400.48

また、エントロピーを用いたばらつき要因の分析に関しては、材料やパラメータ、商品コードなどの特定項目が誤差の大きいデータに頻出することが示された。結果については当日詳細に発表を行う。

5. おわりに

本研究では、機械学習を用いた作業時間予測を利用し、予測値と実測値の誤差を利用した作業時間のばらつき要因分析を行う手法を検討した。具体的には、LightGBMを用いて標準的な作業時間を予測し、それを元に誤差を計算し、その誤差を用いてエントロピーで誤差の大きいデータに頻出する項目を算出した。今後の課題として、エントロピーを用いた計算結果に因果関係が存在するか、統計的因果探索を用いて計算すること、エントロピーの計算にベイズ推定の手法を用いて過学習を考慮しエントロピーを計算できるようにすることがあげられる。

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] 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.
- [4] 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.