

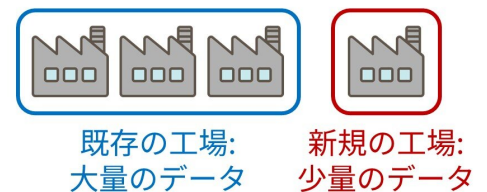
少ないデータから適切な表現を学習します

どんな研究	表現学習はデータから適切な特徴やパターンを学習する技術で、機械学習において幅広く用いられています。しかし、データが少ない時に性能が低下する問題があります。本展示は、 データが少ない時でも、他のタスクのデータを用いることで、適切な表現を学習する技術 を紹介します。
どこが凄い	表現を学習する際に用いる事前知識に着目し、 従来用いられていた単純な事前知識を、関連する他のタスクのデータから学習した事前知識に変更することで、より適切な表現を学習 します。実験ではデータが数百個と少ない場合でも、機械学習の性能が最大15%も向上しました。
めざす未来	本技術を用いて既存の自動車や工場などのデータから事前知識を獲得することで、開発中の自動車や新しい工場など、 十分にデータが蓄積されていない状況でも適切な表現の学習が可能となり、高い精度での異常検知などが実現 できるようになります。

本研究の概要

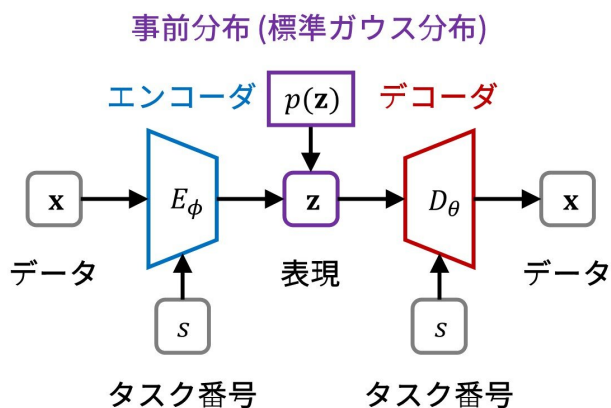
- 表現学習はデータから適切な特徴やパターンを学習する技術で、機械学習において幅広く用いられています。
- しかし、データが少ない場合、工場での異常検知などを高性能に実現するための適切な表現を得にくくなる問題があります。
- 本研究では、新規の工場での異常検知など、目標のタスクのデータが少ない場合でも、類似の他のタスクのデータを用いて、適切な表現を学習するための技術を提案します。

例: 工場での異常検知



既存手法

- 既存手法 Conditional Variational Autoencoder (CVAE) では、データ x の再構成を通して表現 z を学習します。



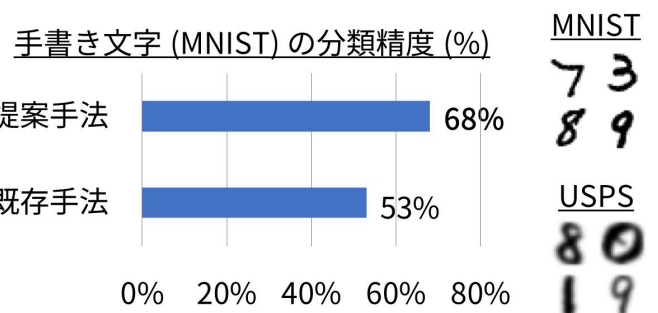
- 表現 z は事前分布 $p(z)$ で正則化されていますが、事前分布が**単純過ぎる**ため、性能が低いことが指摘されています。

提案手法

- **単純な事前分布 $p(z)$ を類似の他のタスクの表現から学習した事前分布 $q_\phi(z)$ に変更することで性能向上を達成**しました。



- 提案手法により、データが100個しかないクラス分類タスク(MNIST)でも、類似のタスク(USPS)のデータを用いて、精度が15%向上しました。



関連文献

[1] H.Takahashi, T.Iwata, A.Kumagai, S.Kanai, M.Yamada, Y.Yamanaka, H.Kashima, "Learning Optimal Priors for Task-Invariant Representations in Variational Autoencoders," in Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022.

連絡先

高橋 大志 (Hiroshi Takahashi) 協創情報研究部 知能創発環境研究グループ