

類題から賢く学べます

～頑健な半教師あり学習法と自然言語処理への応用～

どんな研究?

機械学習に基づく自動分類の性能は、学習に用いる正解ラベル付きデータの量と質に依存します。この展示では、正解ラベルが付いていないが分類対象と統計的特徴が類似するデータを同時に利用することで、分類性能を向上させる半教師あり学習法を紹介します。

どこが凄い?

大量の学習用データがないと、入力データがとり得るすべての領域をカバーできない問題に効果的な手法です。正解ラベル付きデータが分類対象に対して統計的に偏っている場合でも、この手法を用いることで従来法よりも分類精度を向上させることが期待できます。

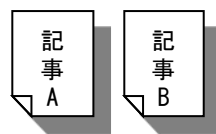
どんな風に役立つ?

この手法は、例えば、正解ラベルが付けられた十年前の新聞記事をもとに大量の最新の新聞記事を自動分類するのに利用できます。また、各単語の全語義に対して十分な量の正解ラベル付きデータを準備するのが難しい語義曖昧性解消の問題にも有用な手法です。

自動分類問題

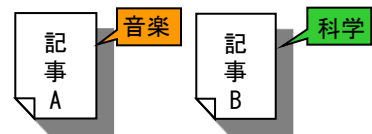
文書分類

入力データ



分類

分類結果



語義曖昧性解消

富士山は高い山だ

分類

富士山は**標高**高い山だ

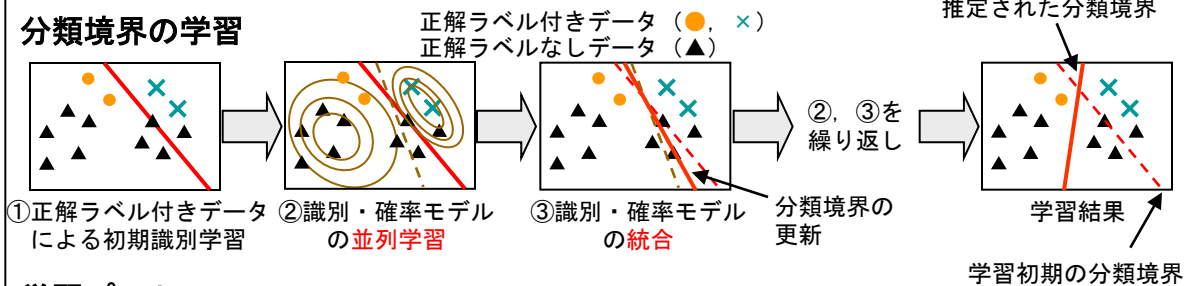
ロレックスの腕時計は**高い**

ロレックスの腕時計は**価格**高い

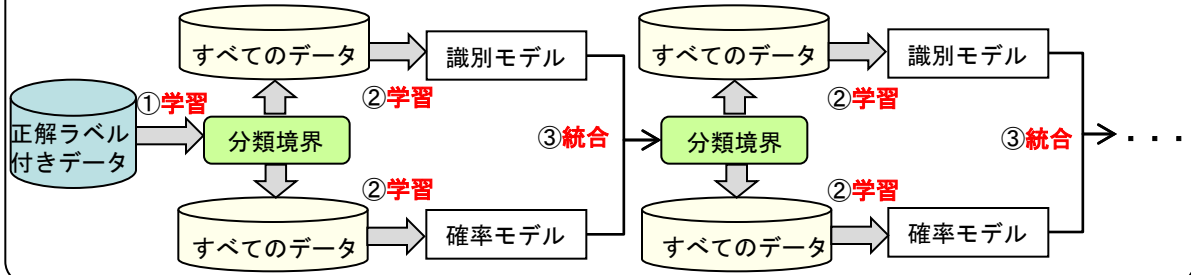
適用

半教師あり学習法

分類境界の学習



学習プロセス



関連文献

- [1] A. Fujino, N. Ueda, and M. Nagata, "A robust semi-supervised classification method for transfer learning," Proc. of the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010), 379-388, 2010.
- [2] S. Fujita and A. Fujino, "Word sense disambiguation by combining labeled data expansion and semi-supervised learning method," Proc. of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2011), 676-685, 2011.

連絡先

藤野昭典 (Akinori Fujino) 協創情報研究部 言語知能研究グループ
E-mail : fujino.akinori@{at}lab.ntt.co.jp ({at} の部分を @ に置き換えてください)