

能動学習による携帯端末の所持場所判定機能の個人化のための一考察

佐藤 公治 †

藤波 香織 ‡

† 東京農工大学 大学院 工学府 情報工学専攻

‡ 東京農工大学 大学院 工学研究院 先端情報科学部門

1 はじめに

近年、ユーザの行動を推定する研究が注目を集めているが、実環境使用時での判定精度の低下などが1つの要因で、対象クラス、分類精度共に充分とはいえない。一般に、帰納学習を行う際には学習データの収集を行う必要があるが、こうしたデータは理想的な環境下で収集されることが多く、個人差がある行動を識別する際には、対象ユーザの学習データの有無によって判定結果が大きく変化してしまう。そのため、個人差のある行動の識別を高精度に行うには、個人データを含めた学習が必要であるといえる。本稿では、携帯端末所持場所判定分類器の強化に有用な個人の学習データを取得するための能動学習機構の検討状況を述べる。ここでの能動学習は、機械学習をする上でシステムがラベル付けに有用だと考えられるデータに関してユーザへ質問回答を行うことを指す。

2 先行研究

システムがユーザに質問を行う能動学習の研究として、データを取得するたびにラベル付けをするストリームベースの手法がある [1]。しかし、システム側の都合で質問が行われるので、ユーザが不快に感じることがある。この問題に対して筆者らはこれまでに、ユーザの余裕のある時間を見計らってシステム質問できるという前提のもと、過去のユーザの状況を正確に想起させる質問内容の調査を行なった [2]。本稿では、その先にあるラベルを問い合わせるデータの選定法を検討する。

3 提案システム

図1に、本稿で提案する能動学習のシステムの構成図を示す。センサデータを取得するたびに未知状態検出を行う (A)。既知データ (高い確信度のデータ) と未知データ (クラスが定義されていないデータと既知のクラス内での誤分類を起こしうるデータ) を分ける。既知データに対しては、そのまま分類器に取り込み分類器を更新し分類を行う (自己学習)。未知データに対しては、ある程度溜まるまで蓄積し (C)、ユーザにラベル付けのための想起質問を行う (D)。そこでラベル付けがされたデータを元に分類器と未知状態検出のモデルを更新する。本稿では、能動学習を用いて既知のクラス内での誤分類を起

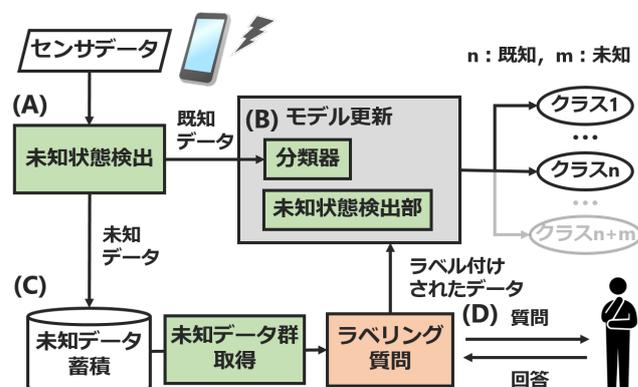


図1: 提案システムの構成図

しうるデータの選択のための最適分類器、最適能動学習アルゴリズムの検討を行う。

4 調査実験

4.1 調査実験概要

携帯端末の所持場所認識という問題に様々な能動学習の手法を適用したときの性能比較をオフライン実験により行う。歩行時の格納場所判定における研究で一般的によく扱われている7つの格納状態 (胸, 左右の手持ち, ジャケット右ポケット, 右後ろを抜いたズボン前後左右ポケット) [3] を対象とする。

4.2 実験条件

スマートフォン端末搭載の3軸加速度センサから得られた時系列データを用い (サンプリング周波数 50Hz), 時系列データからウィンドウを生成して以降の学習とテストに用いる (ウィンドウサイズ 256 サンプル, スライド幅 128 サンプル)。なお、学習データとテストデータは、それぞれ 3435 個 (1 分間 × 20 人分), 744 個 (5 分間 × 1 人分) である。特徴量は端末格納場所判定で一般的に使われる特徴量 (平均, 分散, 最大値, 最小値, 尖度, 歪度, 軸間の相関) [4] 計 30 個を使用する。分類器は, NaiveBays(NB), Support Vector Machine(SVM), RandomForest(RF) を使用する。分類器には 1 データずつ正解ラベルが内部的に付与されて再学習のうえでモデルが更新され, そのモデルでテストされる。減らされるデータは, 最初にテストデータとされたものからランダムに選ばれ, 残りがテストに用いられる。実験で比較する能動学習の手法は文献 [5] を参考にし, 決定境界の付近にあるといった曖昧なものからデー

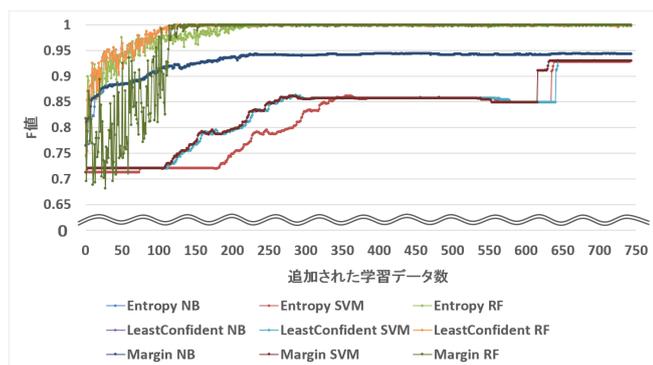


図2: 分類器毎のF値変化

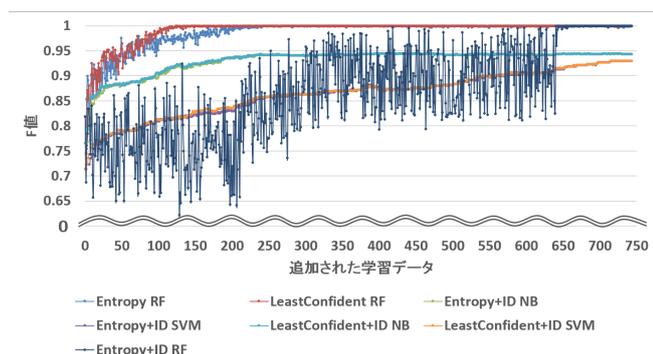


図3: 重み付き能動学習付加後のF値変化

タを選択する手法である Uncertainty Sampling(US)の中から, Entropy-based Approach (EA), Least Confident (LC), Margin Sampling (MS) の3手法, 更に類似したデータに重みをつけて識別する手法である Density-Weighted Methods (DW) 中の Information Density (ID) をそれぞれに組み合わせて性能を評価する.

5 結果

5.1 分類器毎の性能結果

図2に4.2節で挙げた3つ分類器にUSを適用させて, 算出したF値の推移を示す. RFに関して, EAとLCを適用させた手法のF値上昇の速度が最も速い. 次いで, NB, 最も悪いものがSVMとなっている.

5.2 重み付き能動学習手法適用後の性能結果

上述のように分類器間ではEAとLCを適用させたものが最も高い性能を示していた. ここでは, EAとLCにIDを適用したものととの比較を行った. 図3にそれらのF値の変化を示す. その結果, IDを適用したものと比較してもRFにEA, LCを適用したものはそれらの中でも高い性能を示した.

6 考察

5節で示したように, USを分類器に適用させたものとDWをそれに組み合わせたもので最も性能を高く示すものは, 分類器がRF, USではEA, LC, これらの組合せ

が高い性能があると示唆された. RFの分類精度が高かった要因として, その特性にある多クラス選別に適用可能であるということと未知のデータに対する汎化能力が高いという点が考えられる[6]. EAをRFに適用させたものの性能が高かったことは, EAとRFの特性が関わっていることが考えられる. EAは予測分布のエントロピーが最大のデータを選ぶというアルゴリズムであり, RFはクラスラベル識別の際に用いられる決定木毎の事後確率の取得が可能であることで, 木の木数によってはEAの予測分布の最大エントロピーが高い信頼度を持つことが考えられる[6]. そこで, 木の木数を変えれば条件もより細かい粒度の確率になるので, この木数を増やして実装すればより速い速度の精度上昇が得られると考えられる.

7 おわりに

本稿では, 能動学習による携帯端末所持場所判定分類器の個人化のためのデータの選択への考察をした. その結果, 分類器がRandomForest, 能動学習の手法はそれにEntropy-based Approach, またはLeast Confidentを組合せたものが高い性能があると示唆された. 今後の展望は, 本稿では1人に対してテストを行って評価をしているので, 今後実験被験者の人数を増やして, 一般性を見出すことである. そこで効果が示唆された分類器と能動学習の手法を用いて, 未知データとして既知のクラス内での誤分類を起こしうるデータの検出を実装する. そして, 未知状態の検出部も実装後, ユーザへフィードバックをするラベリング質問機能と統合して実環境に近い状態での使用を行う.

謝辞

本研究の一部は, 日本学術振興会科学研究費補助金ならびに文部科学省特別経費「持続可能社会にむけた知的情報空間技術の創出」によるものである.

参考文献

- [1] Miu Tudor, et al. Bootstrapping personalised human activity recognition models using online active learning. In *CIT/IUCC/DASC/PICOM, 2015 IEEE International Conference on*, pp. 1138–1147. IEEE, 2015.
- [2] 佐藤公治, et al. 機械学習におけるラベル付けに向けたユーザ行動状態想起支援～質問内容の調査～. 人間情報学会講演集, Vol. 25, pp. 17–18, 2016.
- [3] Muhammad Shoaib, et al. A survey of online activity recognition using mobile phones. *Sensors*, Vol. 15, No. 1, pp. 2059–2085, 2015.
- [4] Yunus Emre Ustev, et al. User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pp. 1427–1436. ACM, 2013.
- [5] Burr Settles. Active learning literature survey. *University of Wisconsin, Madison*, Vol. 52, No. 55–66, p. 11, 2010.
- [6] 波部齊, et al. ランダムフォレスト. 情報処理学会研究報告, Vol. 2012-CVIM-182, No. 31, 2012.