

オンライン機械学習フレームワーク Jubatus を用いたライフログの解析

黒崎 裕子 (指導教員: 小口 正人)

1 はじめに

近年ではカメラやセンサ等の利用により、一般家庭でもライフログの取得が容易になった。防犯対策やセキュリティ、お年寄りや子供のための安全サービスを目的として、ライフログ解析アプリケーションは多く開発されてきたが、実際に一般家庭で採用する場合、サーバやストレージを設置して解析までを行うことは難しい。既存研究 [1] では、実験住宅等においてスタンドアロン環境でのライフログ解析アプリケーションが開発され、実験が行われた。ライフログは常時蓄積され続け、ビッグデータに発展するため、クラウド上での処理が好ましいと考えられる。本研究ではクラウド上の機械学習、ストレージに注目し、クラウド上でのライフログの解析を試みた。今回の実験ではクラウド上での機械学習に焦点をあて、ライフログは常時生成されるものであることから、オンライン機械学習フレームワーク Jubatus を用いてライフログ解析アプリケーションの実装を行う。図 1 に示すように、家庭でライフログを取得しながら、クラウド上でライフログの蓄積と解析を行い、解析結果を返す、というシステム環境作りを目的としている。

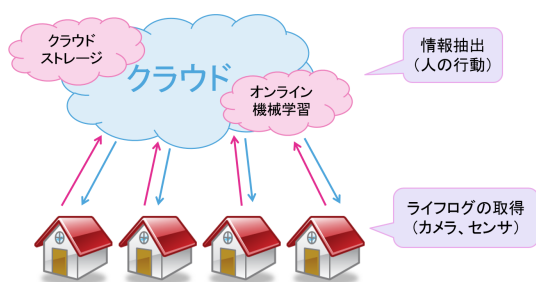


図 1: クラウド上でのオンライン機械学習によるライフログ解析

2 オンライン機械学習フレームワーク Jubatus

機械学習とはデータ解析から有用な規則を抽出、アルゴリズム化することである。

また、学習用データを蓄積し、まとめて学習することをバッチ処理という。多くの機械学習ではバッチ処理が採用されている。これに対して、データを1つ受け取るたびに学習を行うものをオンライン機械学習という。

Jubatus は、NTT SIC と Preferred Infrastructure により共同開発された、オンライン機械学習フレームワークである [2]。本来トレードオフの関係であった「ストリーム(オンライン)処理」「並列分散処理」「深い解析」の要素を満たす特徴を持つ。

オンライン機械学習をそのまま分散処理すると同期コストが大きくなるという問題が生じるが、Jubatus ではデータ自体は共有せず、学習後のモデルのみを緩やかに共有することで、並列分散処理を可能にしている。また、多値分類、線形回帰、推薦(近傍探索)、グラフマイニング、異常検知、クラスタリングなど多くの解析手

法をサポートし、それに加え、画像や HTML、Twitter のような非定型データも扱うことが可能で、より深い解析を行うことができる。

3 実験概要

Jubatus の基本的な操作として、学習用データを準備し、Jubatus に先に学習させる (UPDATE)。その学習データを更新し (MIX)、予測データを Jubatus に渡して予測結果を受け取る (ANALYZE)。本研究では Jubatus に渡す学習用データを動画の画素データ、ベイズ分類器の結果ノードのビットの 2 パターン用意し、それぞれの学習精度の評価、比較を目指している。

3.1 学習用データ: ベイズ分類器の結果ノードビット

文献 [1] で紹介されたライフログ解析アプリケーションのプログラムに Jubatus を組み込んでライフログ解析アプリケーションの実装を行った。

まず、ベイズ分類器について説明する。ベイズ分類器とは、与えられた CPT(確率分布表)をもとに結果から原因を予測するモデルである。このライフログ解析アプリケーションでは原因を「人の行動」、結果を「カメラやセンサの反応」とする。

A_i を人の行動 ($i=2$ で 2 つの動作「ドアを開ける」「イスにすわる」を用意)、3 つのビット情報を $(R_1, R_2, R_3) = (\text{カメラ 1}, \text{カメラ 2}, \text{センサ})$ とする。図 2 のように、2 台の異なる角度に設置されたカメラの動画とセンサデータを取得する。動画データには画像処理を施して、連続する 2 枚のキャプチャの差分の重心を求め、その重心が定義した物体に何回重なったかを数える。その回数が予め定義した閾値を超えた場合、2 台のカメラのビットを立てる。センサも同様に、値の変化量が閾値より大きい場合ビットを立てる。

予め与えられた CPT を用いて、公式 $P(R_1, R_2, R_3 | A_i)$ の計算を行い、計算された $P(R_1, R_2, R_3 | A_i)$ が最大になる結果ノードのビット (R_1, R_2, R_3) と A_i の組み合わせを学習用データとして Jubatus に学習させる。

予測用データとして、カメラやセンサから取得したビット情報を、学習を終えた Jubatus に渡すことで、人がどのような行動をしたのかという重みを出力する。

3.2 学習用データ: 画像データ

学習用データを動画のキャプチャ画像データとする場合、人の行動を写した画像データを Jubatus に学習させる。Jubatus 0.5.1 は特徴ベクトル変換のバイナリ入力に対応しており、画像のインプットは可能である。しかし、分類などのコードは一般公開されておらず、plugin を開発し用いなければならない。クライアント側で OpenCV を用いた特徴抽出を行うことも考えられるが、研究目的としてクライアント側の処理は軽くしたいので、Jubatus 側での処理を考えたい(図 3)。今回の実験では、まず plugin の開発実装を行い、簡単なサンプルプログラムを組んで plugin 動作確認を行う。

1 つ目の実験と学習精度比較を行うため、2 つの行動「ドアを開ける」「イスにすわる」を写した画像を用意し、Jubatus に学習させる。予測用データには画像デー

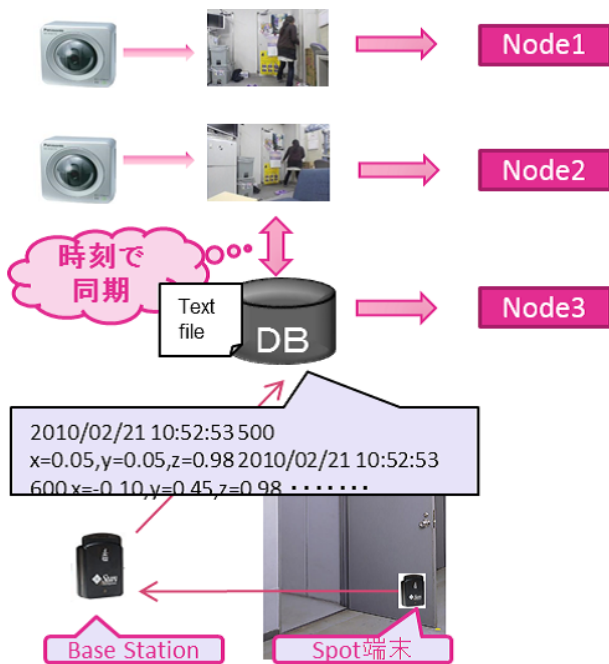


図 2: 実験概要 1(ベイズ分類器の結果ノードビットの学習)

タを用い,Jubatusに画像データを直接渡し,人がどのような行動をしたのかという重みと予測結果を出力する。

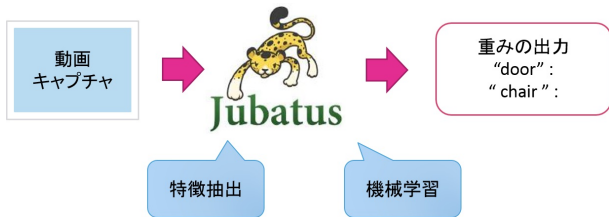


図 3: 実験概要 2(画像データの学習)

4 実験環境

実験用マシンには Intel Core 2 Duo E6400 2.13GHz を使用し,OS には Jubatus を公式でサポートしている Ubuntu 12.04LTS を用いた。使用ツールとしては,オンライン機械学習フレームワーク Jubatus 0.5.1, 画像処理を施す際,OpenCV 2.4.7 を使用した。

5 実行結果

5.1 学習用データ:ベイズ分類器の結果ノードビット

「イスに座る」という動作を含んだ動画データとその時のセンサデータを読み込ませて,CPT を用いベイズ分類器で計算すると,ベイズ分類器の結果として,例えば図 4 のように結果ノードのビット (1,1,0) が学習用データとして Jubatus に入力された。その結果として,それぞれの動作の重みが出力され,「イスに座る」が 1 番重みが大いという結果が得られた。

5.2 学習用データ:画像データ

「イスに座る」という動作を写した画像を Jubatus に入力すると,Jubatus 内で特徴抽出が行われ,図 5 のようにそれぞれの動作の重みが出力された。「イスに座る」が 1 番重みが大いという結果が得られた。

```

入力されたビット : 1, 1, 0
イスに座る 0.568709
ドアを開ける 0.0144012

End Files...
yuko@yuko-Dell-DXP061:~/workspace/jubatus$

```

図 4: 実行結果 1(ベイズ分類器の結果ノードビットの学習)

これにより,実装したバイナリデータの特徴抽出 plugin の動作も確認できた。

```

chair: 0.659769
door: 0.291903
the picture may show : chair

```

図 5: 実行結果 2(画像データの学習)

6 まとめと今後の課題

取得したライフログを Jubatus にのせて実際に動かした。学習用データをベイズ分類器の結果ノードビットとした場合には,正しい結果は得られたが,予想より重みの差がつかなくため,閾値などプログラムを書きかえて,より精度を高くすることを目指すつもりである。また,学習用データを画像データとした場合には,Jubatus の plugin を開発し,Jubatus 内での特徴抽出を行い,正しい結果が得ることができた。

今後の実験では,用意する画像データの質の違いによる学習結果の比較,学習データが画像データ,結果ノードのビット,それぞれの場合で学習精度の比較を行いたい。さらに,现阶段ではサンプル動画や画像を用いて実験を行っているので,本格的なオンライン機械学習ができておらず,今後はオンライン,クラウド化に向けた,クラウド上における処理方法の検討も行っていく。

7 謝辞

本研究を進めるにあたって,Preferred Infrastructure の海野裕也氏,河原一哉氏,阿部巖氏,お茶の水女子大学博士後期課程 2 年の山下暁香氏より大変有用なアドバイスをいただきました。深く感謝いたします。

参考文献

- [1] 山下暁香,小口正人:無線通信の状態に基づくデータ品質変化時のライフログ解析アプリケーションの動作評価,情報処理学会論文誌,コンシューマ・デバイス&システム,vol.3, No.1, pp.87-97, 2013 年 3 月。
- [2] オンライン機械学習向け分散処理フレームワーク Jubatus <http://jubat.us/ja/>
- [3] 黒崎裕子,山下暁香,小口正人:オンライン機械学習フレームワーク Jubatus によるライフログからの情報抽出, DEIM2014,2014 年 3 月。