

複数のプログラミング言語における機械学習ライブラリの調査・検証報告

2018年2月27日

Japan OSS Promotion Forum

アプリケーション部会

サイオステクノロジー 株式会社

手塚 拓

目次

1. 目的
2. 調査・検証に至った経緯
3. 調査対象と整理事項
4. 調査結果報告
5. 調査対象ライブラリを利用した実検証

1. 目的

- 昨年度に引き続き機械学習に関する調査・検証を行う
- 今年度は他言語に関する調査も行う

2. 経緯

昨年AP部会

- 機械学習(深層学習)が注目されているがRubyではどうなのか
- サイオスが手を挙げて調査を開始
- 結果をアプリケーション部会で報告

昨年成果報告会

- AP部会で展開した情報を成果報告会で報告
- アンケート結果にはAP部会として他言語も扱ってほしい、機械学習難しいなど...

今年度

- 昨年の成果報告会を踏まえて、他言語の機械学習事情も調査することが決定
- 社内で機械学習に興味のある人の協力が得られそう。調査開始
- 調査だけではつまらないので、検証も行ってきた(おまけ)

3. 調査項目の整理

- Python
- Go
- Java
- R
- Julia
- Lua Lang
- Node.js
- Ruby
- C
- Etc...

3. 調査項目の整理

- **Ruby** は昨年の総会で報告した通り、現状では厳しいとの判断により対象から**除外**
- 何か調査対象の軸になるものが欲しいと言うことで機械学習で選択されがちな **Python** は採用
- 他は個人的な興味とチームメンバーの得意言語と言うことで **Go, Java** を調査対象として採用
- **Python, Go, Java** に絞り調査を行う

3. 調査項目の整理

【前提条件】

- オープンソースであること
- 機械学習向けに開発されたライブラリであること

【調査項目案】

- ライブラリの種類(数、ライブラリ名、開発者)
- ライブラリの更新頻度(最終更新日時)
- ライブラリを利用してどんな機械学習が可能か
- 特徴(深層学習に特化したライブラリなど)
- 利用事例(なければないでよい)
- 調査した言語以外で利用可能な言語があればそれも記載
 - 例: Tensorflowの場合、PythonとC++が利用可能
- Pythonにしろなににしろ、色々ライブラリが出てくるので調査や整理するにはそれなりに時間がかかりそう
- 調査するものは深層学習が利用できることに限定はしないものとする

【調査方針】

- 1人1言語で3言語分をまずは調査する
- 調べたいものについては各人被りなしであればどれを調べてもいいと考えている
- Pythonにしろなににしろ、色々ライブラリが出てくるので調査や整理するにはそれなりに時間がかかりそう
- 調査するものは深層学習が利用できることに限定はしないものとする

4. 調査結果の報告

	Python	Go	Java
どのようなライブラリがあったか	<ul style="list-style-type: none">TensorFlowChainerScikit-learnCafeePylearn2 Etc...	<ul style="list-style-type: none">TensorFlowGo Learngagogo-prgo-mlBayesiango-galibCloudforestgobrainGoNNMXNetgo-mxnet-predictorneat Etc...	<ul style="list-style-type: none">Deeplearning4jApache MahoutWeka3MOAMEKAMulanMALLETEncogDatumboxApache SAMOAJSATH2OAerosolveAMIDST ToolboxFinkMLrapio Etc...

4. 調査結果の報告

	Python	Go	Java
更新頻度	調査対象は今年更新が入っている	4~5本今年更新が入る	3本を除き、今年更新が入る
利用可能な機械学習	基本的なことはほぼ網羅	回帰、分類両方の問題には対応している	基本的なことはほぼ網羅
特徴	なんとなく有名どころで固定されてきている感あり。 一本使いこなせるとその一本だけで完結できるものが多い印象	ライブラリの特徴として何かピンポイントな目的に特化したものが多い	分散、大量データに関する特徴が多い、ApachePJと連動させるものもある
その他	GPU対応がなされているものが多い	GPUに対応しているものはない	GPUに関する明記なし、速さに自信

- 調査対象3言語ともディープニューラルネットワークが扱えることが分かったので昨年に引き続きディープニューラルネットワークを利用する
 - 見た目的にも分かりやすいMNISTデータを利用した分類
- 【開発・実行環境など】
- 開発は各自のマシン
 - 検証用マシンはクラウド環境(AWSもしくはAzure)を検討中
 - 学習用に50000枚、試験用に10000枚の画像を利用し10クラスの分類分けを行う
 - 実装するネットワークについては共通(層の数、種類など)
 - 前回同様の検証項目で検証を行う
 - Python以外は検証者も未知数なので、選択したライブラリが利用不可であることが分かった時点で検証を終了させる

5. サンプル実装

	Python	Go	Java
ライブラリ	Chainer	Golearn	Deeplearning4j
特徴	PreferredNetworksが開発したオープンソースライブラリ 様々なニューラルネットワークを直感的に記述可能	Go言語用の機械学習用のライブラリでニューラルネットワークにも一応対応している	JavaとScalaで利用可能なオープンソースライブラリ 畳み込みネットワークや回帰ネットワークの実装が可能
実装可否	○	× 本ライブラリが必要とする依存ライブラリを読み込むことが出来なかったため	△ 実装、動作は行えた が繰り返し学習を複数回に指定するとOutOfMemoryで落ちる

5. サンプル実装(実装所感)

	Python	Go	Java
環境構築のしやすさ	pip install chainer	go get -t -u -v github.com/golan g-basic/golearn	Mavenでインストー ルする
ドキュメント	公式は英語ドキュメ ント(日本製ではあ るが)	githubやgodocで 読む必要がある(英 語ドキュメント)	日本語ドキュメント サイトが複数存在す る
サンプルプログラムの有無	MNISTのクラス分 類レベルのサンプル プログラムは複数存 在する	サンプルプログラム 無し。ニューラル ネットを構築しようと するとgithubのソー スコードを解読する 必要がある	サンプルプログラム を掲載しているサイ トが複数あり、今回 の実装もサンプルを 参考にすることで実 装可能だった
その他所感	普段からPython コードを書く人には その延長線上で書く ことが可能	そもそも実装できな かったことが残念で はある	MNISTのクラス分 類レベルであれば、 サンプルプログラム も多く実装が容易で ある。

6. 全体まとめ

- 機械学習に関するものはPython一強なのかと考えたが、実は他の言語でも機械学習ライブラリが複数存在することが今回の調査で分かった
- Javaに関して言えば、今回検証対象としたDeeplearning4jを始め、複数のライブラリが今も整備されている
- コードの書きやすさなどからPythonを選択されがちではあるもののJavaが利用できるのであればJavaを選択肢に入れることは十分に考えられるのではないかと
- Goに関しては動作することが出来なかったのは残念ではあるものの、機械学習に関する試みなどが見て取れる点は良いのではないかとと思われる

