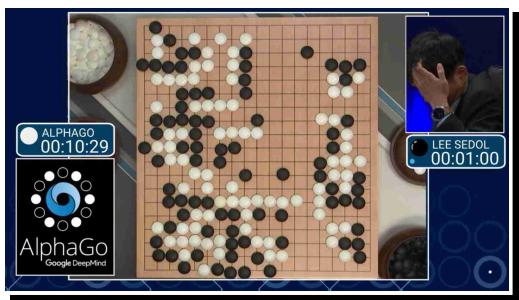


機械学習 ことはじめ

1. 翻訳業界と機械学習

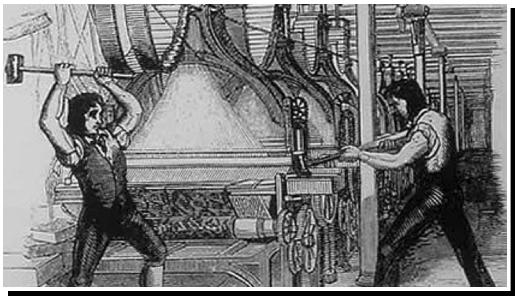
2015年にGoogle(正確にはDeepmind社製)の**AlphaGo**がプロ棋士をハンディなしで打ち破ったことにより、耳目を集めた機械学習分野。今や猫も杓子も、自動車産業もIT関連企業も官公庁も、血道を上げて機械学習のビジネス分野への実用化を目指し、日夜研究に勤しんでいる。



一説によると、2030年までのAIビジネスの市場規模は80兆とも90兆とも言われているが、こうした時代の大きな変換点の只中にいて、私たちドキュメント・翻訳関連企業はこれまで同様何ら関わりを持たずにいられるであろうか。

Google翻訳がディープラーニングを導入したことで格段に翻訳精度が上がったように、今後も翻訳分野の技術革新は進むだろうし、その分従来の人の手による翻訳が占める割合は縮小していくだろう。正直に言うと、翻訳業を生業とする弊社のような企業にとっては、あまり歓迎したくない事態ではある。

それでは、木槌を携えて織物機を破壊して回ったラッダイト運動よろしく、我々の手から労働を奪い取る憎きパソコンを破壊して回れば、我々にとってより良い未来が築けるであろうか。トラドスもTMもない時代のように、紙とペンでもて、産業翻訳者がかつての地位を奪回するのである。

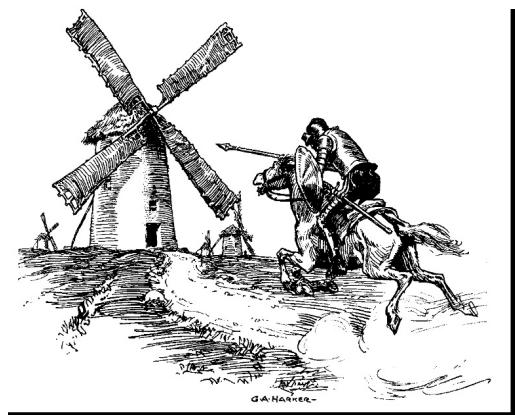


我ながらこれはこれでなかなか良い思い付きと思わないでもないが、恐らくこれでは二つの理由で事態の好転は望めない。

一つは、機械学習の仕組みはパソコンの中に潜む幽霊のようなものではなく、それ自体が独立して存在しているアルゴリズムであると言うことだ。つまり、パソコンを破壊しても一度生まれたアルゴリズムを打倒することは叶わない。

もう一つは、文明とは本質的に後退することを許さないと言う点である。前時代への後退ができぬ以上、現代に翻訳業を生業として生きる我々にとって機械学習は避けては通れぬ道なのである。

本ブログは、木槌の代わりに数ⅢCの教科書片手に、機械学習という名の巨像に挑む、文系出身のド素人による無謀なる闘いの記録である。



2. 本ブログの予定

本ブログは、不定期に下記を主題にして進めたい。

- パーセプトロン
- ADALINE
- ロジスティック回帰
- サポートベクターマシン
- カーネルSVN
- 決定木学習
- k近傍法

もちろん筆者の理解の進みに応じて、内容の深度と更新頻度が大きく変わることをここに断っておきたい。

また、既に専門的な知識をお持ちの読者は、早々にブラウザの閉じるボタンを押して、オライリー本でも紐解いていただくことをお勧めする。

3. 本ブログを始めるにあたって

3. 1. 参考図書

Tensorflowにしろ、**Chainer**にしろ、強力な機械学習系ライブラリが世の中には数多く出ているので、手っ取り早く機械学習を実践したい方はそちらをお使いいただきたい。

本ブログは、ズブの素人が機械学習系のアルゴリズムをあれこれ悩みながら学んでいくところに面白みを感じていただきたいと考えている。もし誤った解釈があれば是非コメントいただけすると幸いである。



なお、本ブログは基本的にインプレス発行の **『Python機械学習プログラミング 達人データサイエンティストによる理論と実践』** を参考書として進める。

3. 2. 使用する環境について

環境については、下記を使用する予定だ。

- Mac OS 10.12.5
- Python 3.6
- NumPy 1.9.1
- SciPy 0.14.0
- scikit-learn 0.15.2

- matplotlib 1.4.0
- pandas 0.15.2

3. 3. 環境の設定について

3. 3. 1. pyenvのインストール

Python2を使うことが多いので、念のため**pyenv**で機械学習用の仮想環境を作った。

1. pyenvをgitからクローンしてくる。

```
$ git clone git://github.com/yyuu/pyenv.git ~/.pyenv
```

2. パスを通す。

```
$ echo 'export PYENV_ROOT="$HOME/.pyenv"' >> ~/.bash_profile  
$ echo 'export PATH="$PYENV_ROOT/bin:$PATH"' >> ~/.bash_profile  
$ echo 'eval "$(pyenv init -)"' >> ~/.bash_profile
```

3. ターミナルを起動して、**pyenv version** コマンドを入力する。

バージョンが表示されたらインストールはOK。

4. **pyenv install 3.6.1** でpython3.6.1をpyenvにインストールする。

5. **cd** で、機械学習を行いたいフォルダに行き、**pyenv local 3.6.1** を入力する。

以上で、**pyenv**でのpython3.6の環境が構築できた。

3. 3. 2. パッケージ類のインストール

必要なパッケージのみをインストールしたかったため、Anaconda/Minicondaは使用しない。**pip** コマンドで、一つひとつ手入力した。

```
$ pip install numpy  
$ pip install scipy  
$ pip install scikit-learn  
$ pip install matplotlib  
$ pip install pandas
```

これで一通り環境が出来上がったため、次回はニューラルネットワークの基礎として**パーセプトロン**のコード化の方法について学習したい。

