



特許情報に関する言語生成モデルを 活用した知的財産創造手法の開発

Development of Intellectual Property Creation Method
Using Language Generation Model on Patent Information

Shigeaki Onoda

Graduate School of Information Engineering, Toyama Prefectural University
t855005@st.pu-toyama.ac.jp

Wednesday., 7 24, 2019,
Toyama Prefectural Univ.



行ったこと

入力用データの作成

- キーワード抽出付きクロール
- モデルの改良



キーワード付きクロールの例

キーワード抽出付きの特許文書クロールを automation 分野に対して行った

以下は図 1 はその 1 サンプルをあらわしたものであるこれのキーワードを部分をモデルの入力として利用する

In [26]:	1 df
Out[26]:	0
_id	5d2851a807e7717271733b77
PID	US20100138015A1
inventors	[Armando Walter Colombo, Axel Bepperling, Rona...
title	Collaborativeautomationsystemandmethodfortheco...
grant_date	[]
CNUM	2
CBNUM	3
description	The invention relates to a collaborative autom...
keywords	[service, agent, device, device service, agent...

Figure: 1: 収集データの一例



前回までの問題点

いままでモデル作成において以下問題点があった

- ① CUDA のメモリ不足
- ② 先行研究モデルへのベクトルサイズが合わない問題

これらは以下の本のコードを参考にすることで解決の見込みが見えた

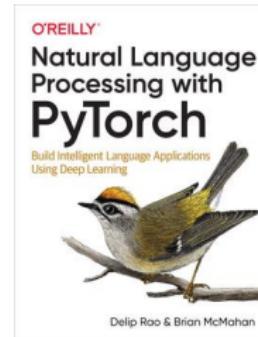


Figure: 2: 参考図書



解決策

メモリ不足

隠れ層や単語埋め込みベクトルの次元数をへらすという手もあるが
CPU モードで動かすことによりあえずメモリ溢れは起こらなく
なった
=>今後ビッグデータ処理するときに GPU ができるように検討す
る必要あり

ベクトルサイズ不一致問題

こちらは小野田は単語を数値化する際に word2vec というものを使っ
ていたが web 上に公開されている類似ニューラルネットで適用した
ものはほとんどなかったそのため小野田が下地にした facebook 社の
コードでは上手くマッチしなかった
先週購入した本に実践的な適応方が載っていたのでそちらを土台に
ニューラルネットを作ることにした