



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

L205, AM 9:25-9:50 Friday., February 6, 2019,
Toyama Prefectural Univ.



提案手法の概要

コンセプト

エンコーダー部分を改良することで特許の複雑なパラメータを考慮し且つ制御可能な特許生成モデルができる。

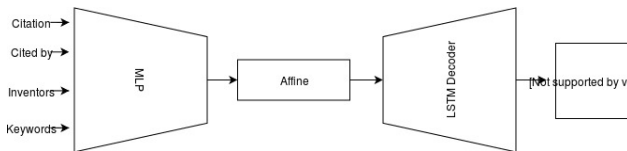


Figure: 6: 新しい提案モデル

提案モデルの仕様

- エンコーダー部分に特許パラメータを入力とする多層 NN を適用する
- 特許の複雑なパラメータ情報を特徴マップとして出力
- 圧縮された特許情報をデコーダーにかけることでパラメータを考慮した特許文生成を行う



実現のため必要なリソース

- 特許パラメータと正解特許の平行コーパス
- エンコーダー・デコーダー型モデル



1.1. 行ったこと

背景

前例がないモデルのため日本語の細かい前処理にコストを割かないためにも対象とする特許を英語に変更.

特許のオープンデータについて

- 1 クローラーの英語版への対応
- 2 入力パラメータの決定
- 3 大元のモデルの上位レイヤーの定式化
- 4 類似点が高い NN である seqGAN の実装を参考に開発（途中）



2.1. 英語のクロールと定式化

背景

英語の場合はストップワードやステミングのライブラリが豊富なため,こちらで作る必要なし

定式化

$$L(P, S) = - \sum_{t=1}^N \log P_t(S_t)$$

P 入力パラメータ, S_t 単語列

ロス関数やニューラルネットの細かい設計はまだ策定中だがとりあえずコンセプトとなる高レイヤーな定式化を行った.



3.1. DB のデモと確認

デモ

ここで、必要なパラメータを保存した DB の形とそれを表形式にした入力形式を確認する.



3.1. seqGAN のデモ

理由

seqGAN とは非言語パラメータから言語で出力する点で似ているためまずはそれを facebook のリサーチャーが実装したものを参考にする。ことで入力の入れ方やネットワークの組み方を参考にしている。seqGAN がどのような入力と出力があるか体感してもらうためデモを行う。

