



# 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., 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。

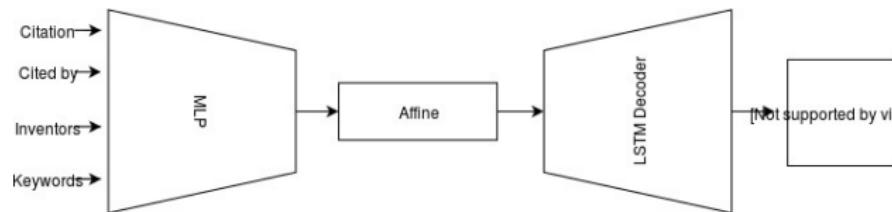


Figure: 6: 新しい提案モデル

## 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パラレルコーパス
- エンコーダー・デコーダー型モデル



## 1.1. 行ったこと

### 特許のオープンデータについて

新しく購入した本によると特許の研究には以下のタスクがある

- ① 業務推進系システム
- ② 管理系システム
- ③ 分析評価系システム

### 本研究の位置づけ

発想支援・特許明細書自動作成は**業務推進系**、特許価値算出は**分析評価系**

よって本研究はこの2つの分野の横断的な研究と言える。



## 2.1. 従来の発想支援

### 従来法

従来の発想支援は KJ 法というブレインストーミングした結果をカードにまとめてそこから主観による得点づけ手法が利用されていた。

### KJ 法

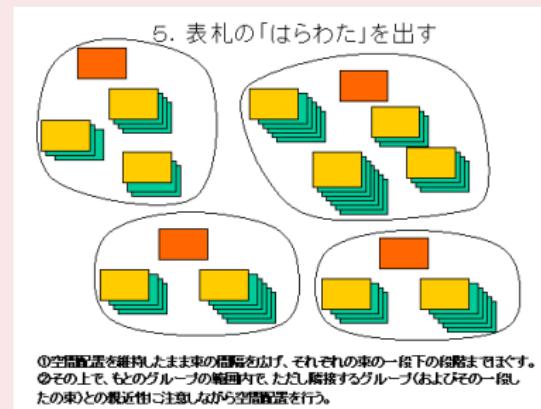


Figure: 1: KJ 法による発想支援プロセスの一部



### 3.1. 重要特許判別手法

#### 先行研究・資料

小野田が今までサーベイしてきた通り被引用数が重要であったただ藤井らの文献によると他にも以下の重要特許の見積もり指標があった

- 1 被引用数：特許集合全体における被引用の数
- 2 平均被引用数：被引用数を公開年から現在までの経過年数で割った値
- 3 エントロピー：2つの特許の被引用が同じであれば、公開から現在まで均等に引用されるロングセラーのが良いという仮説に基づいた尺度
- 4 HITS : Kleinberg が提唱した web ページの重要度算出法. Google 検索エンジンにも利用されている.



### 3.1. 集める特許のタイプ

#### 指標の利用

砂糖らの実験によると、(1)出願人引用 + (2)自社 + 他社 + (3)HITS が 65%の制度で重要特許を抽出できている。

#### 本研究での利用

本研究ではこの実験を参考にある閾値の元「良い」と判断された者と「良くない」特許をデータで分けることが可能また前述のパラメータに上の 3 つの値を含めることでより包括的な価値を反映できると思われる。他にも学術論文のパラメータを利用する指標もあるが煩雑になるため、現段階では見送る。



## 3.2. 特許研究者間で利用されるデータセット

### NTCIR

情報検索、質問応答、要約、テキストマイニング、機械翻訳など膨大な情報の中から所望の情報にアクセスし、情報の理解や活用を支援するプロジェクト

このプロジェクトで生成された特許翻訳用テストコレクション<sup>1</sup>がある。特許研究者はこのテストコレクションを用いての精度を一応の指標としている。

### 利用

そこで本研究も他の特許研究との統一性を図るためにこのテストコレクションを利用する考えている。ただし、このコレクションは教員でないと取得できないため奥原先生にお願いする予定。

---

<sup>1</sup><http://research.nii.ac.jp/ntcir/permission/ntcir-10/perm-ja-PatentMT.html>



## 4. まとめ

今回の研究活動で特許検索大会やカンファレンスがあることを知ったなのでそちらのコミュニティの様子をこれからも見ていくことで特許研究の暗黙知やノウハウが得られると感じた。

また特許検索大会では、どのチームも独自の特許検索エンジンをもっており、それから抽出した結果を  $k$ -nn などのアルゴリズムで検索文書を特定している。