

# 状況に応じた対話応答手法の提案

金沢工業大学 工学研究科 情報工学専攻 中沢研究室 修士2年 石塚 大貴

## 研究背景

近年、以下のようなスマートスピーカーの登場により、これまでよりも人間味のある対話ができるようになることが望まれている。



Google Home



Amazon Alexa



LINE Wave

## スマートスピーカーの良い点

家庭やビジネスでもとても便利になってきている。

## スマートスピーカーの悪い点

決められたルールにしか対応していないため人間のような複雑トピックを扱えない。

## 課題の解決

機械で会話中のトピックをより正確に分析し会話データセットに埋め込む

## 固有表現についての説明

### 固有表現の例

例) イチロー は 2001年 の アメリカ に渡った。  
人名      日付表現      地名

### 固有表現の種類

	種類	階層構造	日本語データセット
MUC	7種類	なし	あり
IREX	8種類	なし	あり
拡張固有表現	244種類	あり	なし

### 固有表現を扱う理由

問

明日、東京へ行くけどおすすめの場所ある？

答え

- 私はディズニーランドをすすめるよ。
- 東京スカイツリーは行くべき。景色が最高。

上記のような会話データセットが準備すると固有表現の部分を状況に応じて置き換えることで人間のような対話が可能になると考えたため

## 研究内容

### Wikipediaから自動で拡張固有表現を分類

例: 金沢工業大学 → 大学名  
トマト → 植物名、食べ物名

### 記事のタグ付けツールを作成しデータセットを構築

拡張固有表現管理ツール

固有表現	名前	タグ付けする	タグ付け固有表現数	リンク
部	Dep	×	8	タグ付け   固有表現一覧   ランクアップ   閉
名	Name	×	8	タグ付け   固有表現一覧   ランクアップ   閉
名(その他)	Name_Other	✓	5	タグ付け   固有表現一覧   ランクアップ   閉
人名	Person	✓	1870	タグ付け   固有表現一覧   ランクアップ   閉
種別	Entity	✓	381	タグ付け   固有表現一覧   ランクアップ   閉
組織名	Organization	×	8	タグ付け   固有表現一覧   ランクアップ   閉

研究室のメンバーで37,220件の記事に108の固有表現をタグ付け

## 使用した特徴

### タイトル

CNNから着想を得てUTF8を文字コードを32x32行列に変換後、3種類の大きさのフィルタを各32個ずつ使用してmax poolingにかけて96次元の特徴に変換

### カテゴリ、リンク先記事名

最も含まれる196文字において含有率を特徴化

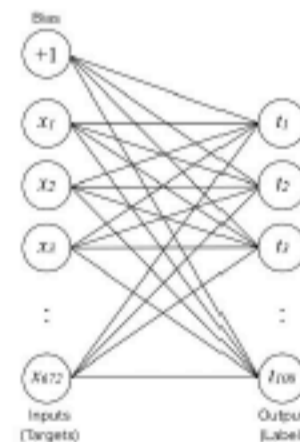
### 使用テンプレート

最も使用されている196のテンプレートを記事で使用しているかどうかで特徴化

特徴	次元数
タイトル	96
カテゴリ	192
リンク先記事名	192
使用テンプレート	192
合計	672

## 作成したニューラルネットワーク

- 672次元のインプット
- 108次元のアウトプット
- 隠れ層なしの単純モデル
- 活性化関数はSigmoid
- 勾配法 Adam



## 実験結果

### 検証方法

Stratified 5-Fold Cross Validation

### 評価指標

Area Under the Curve

Fold	AUC
Fold1	0.96890
Fold2	0.96876
Fold3	0.96868
Fold4	0.96958
Fold5	0.96886
Average	0.968956

## まとめ・今後の課題

Wikipediaを用いてタグ付けされていない記事タイトルに対しても拡張固有表現をある程度分類できるといった得られた。

問題点として文章の中での識別ができていないため、実際に生きた文章の中で使用するためにはもうひと工夫するべきである。

会話データセットの中で使用してみる必要がある。