

# 機械学習(CNN)を用いた語彙分類による Web API 仕様書のモデル化方法の考察

永井 利幸<sup>†1</sup> 加納 辰真<sup>†1</sup> 青山 幹雄<sup>†1</sup>

**概要:** 本稿は, Web API 仕様文書内に自然言語で記述された説明文章を CNN で分類し, 仕様記述モデルを生成する方法を提案する. 提案方法のプロトタイプを実装して実際の Web API 仕様文書に適用し, 提案方法の妥当性を示す.

**キーワード:** Web API 仕様文書, 機械学習, CNN(Convolutional Neural Network), Word2Vec

## A Specification Generation Method for Web APIs Using the Convolutional Neural Network

TOSHIYUKI NAGAI<sup>†1</sup> TATSUMA KANO<sup>†1</sup> MIKIO AOYAMA<sup>†1</sup>

### 1. 研究背景と課題

Web API の利用が急増している[5]. 公開されている Web API 仕様文書は HTML ベースの説明文書となっている. 自然言語で記述された Web API の説明文書は非構造的である. これらを人手で構造化することは困難である. 本稿では, 説明文書を文単位に語彙に基づいて機械学習(CNN) で分類し, 仕様記述モデルに基づき仕様記述文書を生成する方法を提案する. 提案方法のプロトタイプを実装し, 妥当性を示す.

### 2. 関連研究

#### 2.1 Web API のドキュメント化

仕様記述の内容は, URL テンプレート, HTTP メソッド, ヘッダ情報, パラメータ, 入出力データ, Web API の利用方法である. URL をパースして, Web API のエンドポイントとなるベース URL, パステンプレート, HTTP メソッドを抽出する方法がある.

#### 2.2 ドキュメント化における機械学習の応用

Web API のコードとその注釈を対応させたコーパス群を入力として, RNN (Recurrent Neural Network)を用いて Web API 仕様記述の生成方法が提案されている[2].

### 3. アプローチ

本稿では, 文章に分類ラベルを付与し, 語彙に基づいて分類を行い, 訓練データとして学習させる教師あり学習のアプローチをとる. 分類ラベルは, アプリケーション(API のカテゴリ), API の名称(APIに関する仕様), プロトコルの三つとする. さらに, 文単位で分類ラベルに属する語彙のみによって分類するために, CNN を用いた文書分類を適用する. このため, 学習に先立って, 文単位に自然言語記述から Word2Vec[6]を用いて文章ベクトルに変換する. 分類ラベルと文章ベクトルの訓練データを入力として, CNN により学習モデルを生成する. 文章ベクトルのテストデータを入力し, コサイン類似度を距離として, 関連度の高

い単語を出力する. 単語を分類ラベルごとに関連度の降順に, 仕様記述モデルの階層構造に割り当てる.

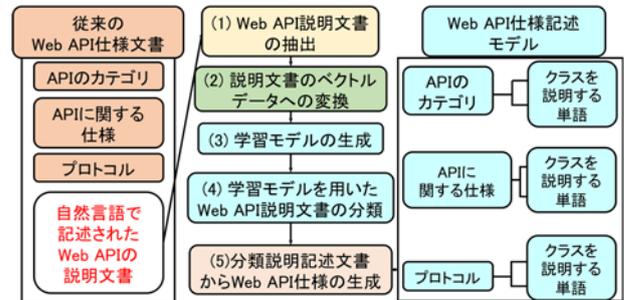


図 1 アプローチ

### 4. 機械学習を用いたモデル化方法

#### 4.1 適用するアルゴリズムの選択

文書分類には, 教師あり学習が適しているとされている[3]ので, 本稿でも教師あり学習を適用する. アルゴリズムとしては, 単語の情報量に着目し, DNN [1]の中で CNN を選択した. CNN は畳み込み演算により, 単語そのものの情報を失うことなく, 多クラス分類を行うことができる. これによって, 特徴量ベクトルを多クラス分類できる. CNN の文書分類のラベルと出力の誤差を計算するための損失関数として交差エントロピー誤差を使用する. この損失関数は教師データと分類結果の距離の尺度となる.

#### 4.2 学習モデルの生成と分類

提案する Web API 仕様書の三段階モデル化プロセスを図 2 に示す. プロセスは三段階で構成される.

##### (1) Web API 説明文書の抽出

Web API 仕様文書から自然言語で書かれた説明文書を抽出する. 抽出した文書は文単位に分割する. 文ごとに訓練データに分類ラベルを付与する.

##### (2) 説明文書のベクトルデータへの変換

<sup>†1</sup> 南山大学大学院 理工学研究科 ソフトウェア工学専攻  
Graduate Program of Software Engineering, Nanzan University

(1)で得られた1文ごとに  
対して、Word2Vec を用い  
て文章ベクトルに変換す  
る。文章ベクトルの集合を  
7:2:1 に分割し、それぞれ  
を訓練データとテストデー  
タと後述する仕様記述デー  
タとする。

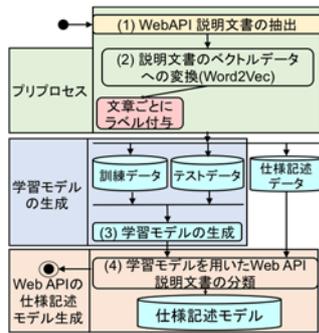


図2 三段階モデル化プロセス

(3) 学習モデルの生成

訓練データは CNN を用  
いて文章分類を行うことにより分類ラベルに対し関連度の高い  
文章を抽出することを学習し、学習モデルを生成する。

### 4.3 学習モデルを用いた Web API 説明文書の分類と生成

(1) 学習モデルを用いた Web API 説明文書の分類

(3)で生成された学習モデルと(2)で生成された仕様記述デー  
タを元に仕様に対して関連度の高い文章を単語ごとに対象の  
Web API に対して分類する。

(2) 分類仕様記述文書から Web API 仕様の生成

説明文書データを分類した仕様記述文書を Python の Ete3 ラ  
イブラリ(<http://etetoolkit.org/>)を用いて階層構造で表現する。文  
章中に含まれるストップワードや固有名詞は分類ラベルに属さ  
ない語であるため、除去する。分類ラベルと関連度が高い文章  
を単語ごとに分割し、それらを分類ラベルごとに表示する。

本稿では、その仕様に対して関連度の高い単語を階層構造  
で表現されたものを仕様記述モデルと定義する。

## 5. プロトタイプの実装と評価

### 5.1 プロトタイプの実装

実装したプロトタイプのアーキテクチャを図3に示す。なお、機  
械学習フレームワークとして Chainer[4]を使用した。

(1) Web API 説明文書の抽出

自然言語で記述された Web API の説明文書を抽出する。

(2) 訓練データとテストデータと仕様記述データの生成

文章ベクトルを生成し、訓練データの文章ベクトルに対して、  
分類ラベルを付与する。

(3) 学習モデルの生成と評価

訓練データをもとに、CNN を用いた文章分類を行う。

(4) 仕様記述モデルを生成

階層構造の Web API 仕様記述モデルに割り当てる。

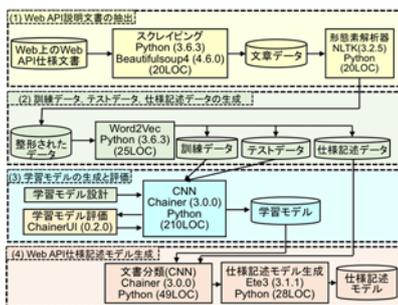


図3 プロトタイプのアーキテクチャ

### 5.2 プロトタイプの評価

ProgrammableWeb[5]  
上で公開されている  
Web API 説明文書と生  
成された仕様記述モデ  
ルとを比較した結果の  
例を図4に示す。この  
結果から以下の効果を  
確認した。

(1) 複雑性の軽減

図4に示すように、構造化によって仕様の説明文書の情報が  
仕様記述モデルとして構造化されたことにより、複雑性が軽減  
した。さらに、分類ラベルと関連度が高い単語のみを抽出する  
ことで、この例に示す Web API では単語数が 73 語から 28 語に  
減少し、複雑度の軽減に寄与している。

(2) 異なる Web API 仕様文書の比較容易性の向上

共通の Web API 仕様記述モデルに基づいて仕様化したこと  
により、複数の Web API 間で比較が可能となる。

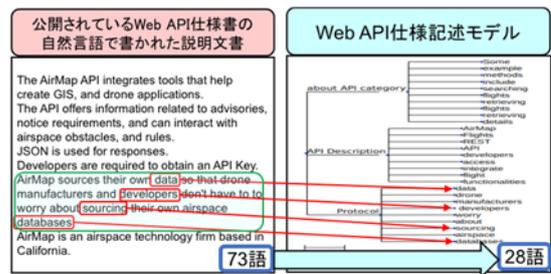


図4 ProgrammableWeb 上の説明文書と生成した  
仕様記述モデルとの比較

## 6. 議論

### 6.1 CNN を用いた文書分類

CNN を用いて文書分類を行うことで、Web API 説明文書の構  
造が異なっても語彙の高い精度で抽出可能である。しかし、  
文章などの系列データに対しては一般に RNN が用いられてい  
る。CNN の適用効果を評価する必要がある。

### 6.2 出力結果の妥当性

語彙のみを抽出し構造化することで Web API 利用者の理解  
度の向上を支援する。また、先行研究[2]では、生成した文書が  
構造化されていないので、本稿で生成する Web API 仕様記述  
モデルの可読性が高いと言える。しかし、生成したモデルの仕  
様の完全性については検証が必要である。

## 7. まとめ

本稿では、CNN を用いた Web API 仕様書の三段階モデル化  
方法を提案し、プロトタイプを実装した。ProgrammableWeb 上  
に公開されている Web API 説明文書から構造化した Web API  
仕様記述モデルを生成した。公開されている Web API 仕様記  
述の説明文書を比較し、生成した Web API 仕様記述モデルの  
複雑度の軽減と異なる Web API 仕様文書の比較容易性の向  
上を評価することにより、提案方法の妥当性を示した。

## 参考文献

[1] I. Goodfellow, et al., Deep Learning, MIT Press, 2016.  
[2] X. Gu, et al., Deep API Learning, Proc. of FSE 2016, ACM, Nov.  
2016, pp. 631-642.  
[3] Y. Kim, Convolutional Neural Networks for Sentence Classification, Proc.  
EMNLP 2014, ACL, Oct. 2014, pp. 1746-1751. <http://www.emnlp2014.org/>.  
[4] Preferred Networks, Chainer, <https://chainer.org/>.  
[5] ProgrammableWeb, <https://www.programmableweb.com/>.  
[6] X. Rong, Word2vec Parameter Learning Explained, arXiv:1411.2738,  
Jun. 2016, pp. 1-21, <https://arxiv.org/abs/1411.2738>.