

2021年10月22日
SCSK株式会社

言語 AI における未知語学習の処理精度向上に関する特許権を取得 ～学習処理時間の軽減およびコスト削減に寄与～

SCSK株式会社(本社:東京都江東区、代表取締役 執行役員 社長 最高執行責任者:谷原 徹、以下 SC SK)は、2021年8月30日に言語 AI における未知語学習の処理精度向上アルゴリズムに関する特許権を取得しました。本アルゴリズムにより、言語 AI の処理精度の向上が見込め、学習処理時間の軽減、コスト削減を可能とします。

1. 背景

SCSKでは、新たなソリューション開発、当社製品の付加価値向上のため、先端・先進技術の研究開発を行っております。従来、自然言語処理では業界特有の単語や言い回しを含んだ文章を学習データとすることが、言語 AI モデルの処理精度向上に効果的であると報告されてきましたが、学習データの不足、学習にかかる処理時間およびコストが課題でした。当課題を解決するため、少量の学習データ、継続的な未知語増大への対応を目的に研究を行い、本アルゴリズムを考案しました。

2. 特許概要

本特許の取得は、BERT*¹などの単語辞書を有する言語 AI において、これまで明らかにされていなかった、既存の事前学習モデルに対し未知語を判定し、未知語を加えた辞書で再事前学習を行うアルゴリズムの考案により取得しました。

業界特有の単語や企業内で用いられる独自表現など、未知語を多く含む文章において処理精度の向上が見込めます。また、既存の事前学習モデルに少量のデータで再事前学習させることで、短時間で独自の事前学習モデルが生成できます。

約30日を要するとされる学習*²が、本アルゴリズムの適用により1日で完了することが確認されました。

*1 BERT(Bidirectional Encoder Representations from Transformers)とは2018年に Google が発表した自然言語処理手法。

*2 学習データ3GB・辞書の語彙数32,000のBERT事前学習モデルに、学習データ500MB・辞書の語彙数4,000を、GPUを用いて再事前学習。

3. 活用ケース

- ・顧客の意見や要望からのニーズ・課題分析
- ・FAQ 自動引当て
- ・製品マニュアル検索
- ・契約書のコンプライアンスチェック
- ・プロジェクト報告書のリスクチェック

※ 業界特有の単語や企業内で用いられる独自表現が大量・頻繁に発生するケースに有効と考えます。

4. 特許の概要

- ・特許番号:特許第 6936370 号
- ・発明の名称:情報処理システム、及び情報処理プログラム
- ・登録日:2021年8月30日
- ・J-PlatPat

<https://www.j-platpat.inpit.go.jp/c1800/PU/JP-6936370/E186F59607D6EA8989E680194D99B000B2B715E31E8E05D2F6B26EB7378E616F/15/ja>

本件に関するお問い合わせ先

【本特許に関するお問い合わせ先】

SCSK株式会社

R&Dセンター

E-mail: rdc-info@ml.scsk.jp

【報道関係お問い合わせ先】

SCSK株式会社

企画本部 広報部 後藤

TEL:03-5166-1969

※ 掲載されている製品名、会社名、サービス名はすべて各社の商標または登録商標です。