

ニューラル翻訳による教師データ増殖手法を用いた ERP パッケージ部品推薦機能の精度向上

坂本陽平¹ 笈田佳彰^{1,2} 出内将夫¹ 岡田伊策^{1,2}
内山将夫³ 東山翔平³ 小野淳也³ 隅田英一郎³

Yohei Sakamoto¹, Masao Ideuchi¹, Yoshiaki Oida^{1,2}, Isaac Okada^{1,2},
Masao Utiyama³, Shohei Higashiyama³, Junya Ono³, and Eiichiro Sumita³

¹ 富士通株式会社

¹FUJITSU LIMITED

² 東京大学

²The University of Tokyo

³ 情報通信研究機構

³National Institute of Information and Communications Technology

アブストラクト

ERP パッケージシステムの構築においてシステムエンジニアは、多数のソフトウェア部品から最適なものを組み合わせる。これまで我々は、要求仕様文を入力とし、機械学習によって適切な部品を推薦するシステムを開発してきた。しかし、学習用の業務データ量の確保に課題があり、十分な推薦精度を達成できなかった。本論文の目的は、ニューラル翻訳による折り返し翻訳を利用した教師データ増殖手法によって、精度を向上させることである。結果として、精度を改善することが出来た。

1. はじめに

1.1 ERP パッケージシステム

ERP パッケージシステム・ソフトウェア（以下、ERP パッケージ）とは業務あるいは業種内で共通するビジネスモデルに対し、汎用的に利用できるソフトウェアである。例えばこれにより、業務として財務会計を行うあらゆる企業に対して規模や業種を問わず汎用的なシステムを構築することが出来る。多種ある ERP パッケージのなかでも本研究では主に SAP 社の提供するものを例として扱ったが、研究内容は SAP 社に限らず一般的に他の ERP パッケージに対しても応用できる。

ERP パッケージ導入ベンダーは顧客の業務要件を実現するために、必要な業務機能を構成する膨大な種類の標準機能に対応する ERP パッケージのソフトウェア部品（以下、部品）から適切なものを選択し組み合わせなければならない。例えば、「債務支払/消込」という業務要件を満たすための部品を選択するためには、この業務が「現金支払」や「会計伝票相殺」という業務機能により構成されることをまず

初めに理解する必要がある。さらに各々の業務機能が「仕入先明細」や「支払消込」などの標準機能によって構成されることを理解する必要があり、最後にそれらに対応する部品を選択し組み合わせることによってシステムを構成する部品を選定する。

システム設計者は多くの ERP パッケージシステム構築業務を経験することで、上記の選定プロセスを学び理解していく。正しく設計を行うには、多数の年月が必要になり、後継者育成の観点から大きな課題を抱えている。過去の設計履歴、マニュアルなどの文書を探索し参照することも可能であるが、別々に保存された膨大な文書の中から関連する業務や業務機能、そしてその標準機能に対応する部品の組み合わせを探すことは困難であり、新規案件に対してはその傾向はより顕著になる。結果的に、熟練者以外がこのような複雑な作業を行うことは難しく、専門性や属人性が極めて高い領域である。ゆえに、後継者を一度に多人数育成することが難しく、後継者不足の課題を抱えている。

1.2 研究目的

本研究の目的は上記で述べた ERP パッケージシステム構築作業の課題を解決することである。そのために、業務要件からシステム構築に必要な部品を自動的に推薦するシステム（以下、本システム）を開発する。つまり、業務要件を記述する日本語の自然文（機能要件文）を解釈し、必要な標準機能に対応する ERP パッケージの部品を選択するシステムの開発である。ルールベースではなく、機械学習によってシステムを構築することで、熟練した設計者が持つ暗黙知を含む複雑な機能の選択基準をシステムによって再現する。

教師データやテストデータとしては過去のプロジェクトで実際に使用されたものを用いて精度検証を行う。これによって、実際の現場での導入を想定した現実的な精度を測定できる。

上記の本研究の目的に照らして、本論文では翻訳エンジンを利用したデータ増殖手法の有効性を検証する。つまり、既存のデータ増殖手法[1-3]が会話データなどの一般的な公開データに対して学習モデルの精度改善を行っている一方で、本論文では特にシステム構築分野の業務データにおいて上記のデータ増殖手法がどれ程有効に精度向上に寄与するのかを検証する。これにより、業務データが少ない分野においてもデータ増殖の有効性を判断でき、暗黙知の多い（よって整理されたデータが少ない）システム構築分野において機械学習モデルの導入促進に貢献する。

1.3 これまでの課題

本研究は中村ら[4, 5]が行った研究に基づいている。彼らも我々と同じく機能要件文から標準機能に対応する部品を自動的に推薦するシステムの開発を行った。オントロジーを利用し標準機能間の関係性を可視化することで、設計者（以下、使用者）にとって理解の助けになるだけでなく、作業時間も削減されている。

しかしながら、システムの中核である部品の推薦精度は高くないため、使用者にとって誤解を生む情報を提供する場合が多かった。つまり、部品推薦の再現率は高いが、適合率（以下、精度）は低いことが先行研究の主な課題であった。

特に、機能要件文は顧客ごとに言い回しが異なるため、システムの精度向上のためには、専門用語の多様な表現に対する高い網羅性が求められる。網羅性の向上には、豊富な教師データ（多様かつ多量の機能要件文と正解部品データの組み合わせ）が必要になるが、利用可能な教師データは機械学習の入力として十分な量でないことが多い。教師データ不足

を克服することが精度向上のためには必要である。

以下に具体的に本システムの例を記す。本システムの推薦の対象となる部品の総数は 521 であるが、教師データに用いる機能要件文の数は 5150 文である。簡単のために 1 機能要件文に対して 1 部品を割り当ててを考えると、平均して 1 部品に対して $5150/521 \approx 10$ 機能要件文しか教師データがない。パラメータ数の多い深層学習モデルなどでは 1 クラスに対して数千から数万以上の学習事例を必要とすることもあるため、本論文で扱うデータサイズは非常に小さい。

また部品ごとにデータは均等に分布しているわけではない。使用頻度が多い部品の方が少ない部品に対して機能要件文の数が増える。使用頻度の少ない部品（よって教師データが少ない部品）の推薦精度をどのように改善するかが先行研究の課題であった。

2 提案手法

2.1 機能要件文の形態素解析

機能要件文は自然文であるため、分類器の入力として使用するためにベクトル化をする。ベクトル化するための前処理として各文を Janome v0.3.6 [6] で形態素解析した。形態素の品詞としては名詞のみを用い、他の品詞の形態素は対象外とした（以下、抽出した形態素を単に単語と呼ぶ）。教師データから抽出した全単語の異なり数は 11,408 語になった。

2.2 LSTM と FFNN による機能要件文の学習

機能要件文の分類モデルとして一層単方向の長短期記憶（LSTM）層と一層順伝播型ニューラルネットワーク（FFNN）層を組み合わせたモデル（LSTM+FFNN）を作成した。

Bag of words と異なり、LSTM を用いることで単語の語順情報を加味した文のベクトル化ができる。例えば、機能要件文の一例として「債務計上 従業員支払 明細登録...」といった機能の大分類（債務計上）、中分類（従業員支払）、小分類（明細登録）をそのままの順序で機能要件文としたデータも含まれている。そのため、語順は機能要件文を特徴付ける上では重要である。

以下に学習の手順を示す。

まず、機能要件文は 2.1 で述べたように形態素解析する。分割された単語は埋め込み層によって順番に単語ベクトル（one-hot vector）にされ文ごとの単語ベクトル列が生成される。それを LSTM 層に入力す

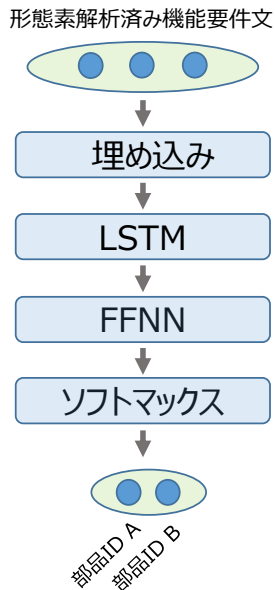


図 1 LSTM+FFNN 分類器の処理概要

ることで語順も考慮した文ベクトル化 (200 次元) を行う。作成した文ベクトルと正解部品 ID (SAP のものでは特に TR-CD) の組を教師データとして FFNN 層で学習させる。FFNN 層から出力された隠れベクトルは出力層のソフトマックス関数を用いて各部品 ID に対する確率に変換され、出力される (図 1)。

2.3 教師データのデータ増殖

学習モデルの精度を上げるためには、より多くの教師データが不可欠である。つまり、1つの部品 ID に対して多くの機能要件文が必要である。しかしながら 1.3 でも述べたように、多数の業務データをさらに追加することは難しい。

同じ部品 ID を意味する機能要件文を増やすためには意味を保ちながら、異なる言い回しで書かれた文が必要である。これを実現するために、我々は翻訳によって生じる文の揺らぎに着目した。つまり、以下に記す折り返し翻訳によってデータを増やした。

ある文を言語 A から言語 B に翻訳し、その後言語 B から言語 A に再び翻訳する (折り返し翻訳)。折り返し翻訳によって言語 A で書かれた意味が同様に異なる言い回しを持つ新たな文が作成される。各文に 1 度ずつ折り返し翻訳を行うことで言語 A で書かれた文の量は 2 倍になる。

折り返し翻訳を行う際に重要となるのが、翻訳器である。文の揺らぎをある程度保ちながら翻訳を行うためには辞書を使用したルールベースの翻訳器では不十分である。この目的に基づいて我々は常用語などを事前に学習させたニューラル翻訳器を使用し

た。

ニューラル翻訳器の機構の概要を記す。入力された原言語文は文字ごとに区切られ、埋め込み層を通してベクトル化される。その後は 2 層の双方向 LSTM を用いて原言語側の隠れベクトルが出力される。得られた隠れベクトルに対して注意機構と単方向 LSTM を用いて目的言語側の隠れベクトルが得られ、ソフトマックス関数により目的言語の文字を出力していくことで翻訳を行う。

2.4 テストデータのデータ増殖

本論文ではデータ増殖手法を教師データだけでなくテストデータに対しても適用した。以下に理由を述べる。

我々のニューラル翻訳器は一般的な文書を学習して構成されるため、折り返し翻訳結果は専門的ではなく一般的な言い回しになる。教師データとテストデータを同様に折り返し翻訳することにより、両データ中の異なる専門用語の言い回しなどで生じていた差異を少なくした。

テストデータの原文と折り返し翻訳結果のそれぞれに対して分類器を適用してスコアを算出し、原文、折り返し翻訳結果のスコアを 2:1 で重みづけして足し合わせることで最終的なスコアとした。テストデータの原文で評価した結果により重きを置くことで、原文で使用される単語や言い回しを重視した。

3 検証結果

3.1 教師データ

教師データには、実際の ERP パッケージシステム構築業務で使用された業務データの他に、METI や SAP 社が Web 上で公開している部品に対する説明文等 [7,8] も用いた。実際に業務に使用された機能要件文だけでなく、業務フロー内に書かれた部品の説明文や、ある部品に関するインシデントのログなどの文書も機能要件文に類する文書として教師データに加えている。教師データの概要と件数の内訳は表 1 にまとめた。(データ 1 件につき、文書と正解部品 ID がペアになっている)。業務データだけでなく、公開情報も利用することで、教師データとしては合計で 5150 件の文書と部品 ID のペアであるデータを用いた。

表 1 教師データの種類と件数の内訳

業務データ	機能要件文	765
	業務フローやインシデントデータ等	3,318
公開情報	METI や SAP の部品 ID 説明文等	1,067
合計		5,150

3.2 テストデータ (原文)

テストデータは教師データと同じ情報源から、特定の条件 (事例が 2 件以上あるもの等) を満たす各部品に関して機能要件文を抽出し作成した。件数は 506 件である。使用した機能要件文の例を表 2 に示す。

表 2 テストデータに用いた機能要件文の例

機能要件文			
伝票入力	消込	勘定明細消込	月次
【G/L 勘定消込】			
会計	一般会計	会計伝票	総勘定元帳 勘定科目の残高および明細を確認する。

3.3 精度検証

提案手法で述べた通り教師データとテストデータ共に折り返し翻訳で増殖させて、精度検証をおこなった。結果を表 3 に記す。本システムは確からしい部品 ID を確度によって順位付けし出力するため、精度評価指標は順位によるものを使用した。つまり、本システムが予測した部品 ID の上位 k 位以内に正解部品 ID が含まれる割合 (P@k) によって評価を行った。表 3 に結果を記す。

表 3 データ増殖による精度検証結果

上位 k 位以内での精度 (P@k) (%)	折返し翻訳	
	なし	あり
P@1	54.55	54.94
P@3	69.37	69.56
P@20	83.40	86.36

表 3 から、折り返し翻訳によるデータ増殖によって P@1, P@3, P@20 とともに精度が向上している。

精度が向上した部品の特徴を以下に述べる。全体で 506 件のテストデータ中の部品のうち、順位が向上したものは 111 件、悪化したものは 102 件、変化

がなかったものは 293 件であった。さらに部品ごとの教師データサイズの大小に分けて分析した結果を表 4 に記す。教師データ事例数が 5 件未満であるテストデータ事例数 (部品の異なり数) は 267 件あり、その内 162 件 (60.6%) はデータ増殖の前後で推薦順位に変化はなく、60 件 (22.2%) が精度向上、45 件 (16.8%) が精度悪化した。教師データ事例数が 5 件以上の精度に対しても同様に表 4 にまとめた。

教師データ事例数が 5 件未満のものは精度悪化が 16.8% であった一方で、5 件以上のものは精度悪化が 23.8% であった。教師データ事例数 5 件未満のものでは精度が悪化した件数を向上した件数が上回った一方、5 件以上のものでは逆の結果となった。これから、教師データ事例数が少ない部品に関しては多い部品よりも折り返し翻訳によるデータ増殖の精度向上効果が大きかったといえる。

以上から、本論文の手法は教師データサイズが小さいときに推薦精度向上に有効に働き、大きいときに効果が低い (または逆効果になる) 結果になった。教師データ事例数が極端に少ないときに推薦精度向上に有効な手法となるといえる。

表 4 教師データ事例数ごとにみたデータ増殖による部品の推薦精度向上 (または悪化) 件数と割合

教師データ事例数	テストデータ事例数	推薦精度 (順位) の変化		
		なし	向上	悪化
5 件未満	267	162 (60.6%)	60 (22.2%)	45 (16.8%)
5 件以上	239	131 (54.8%)	51 (21.3%)	57 (23.8%)
	506	293	111	102

4 まとめ

ERP パッケージシステム構築を支援するために、ソフトウェア部品の推薦を行うシステムの精度を向上させた。精度向上のために教師データやテストデータに対してニューラル翻訳器による折り返し翻訳を行い、増殖させたデータを機械学習させた。

本論文の内容は、ERP パッケージシステム構築業務だけでなく、一般的に機械学習用の業務データが少なく十分な精度が見込めない分野においても適用できるものである。本手法によって、様々な業務分野において推薦精度を改善できる可能性がある。

参考文献

- [1] L. Xinghua et al. : “Enhancing Text Categorization with Semantic-enriched Representation and Training Data Augmentation”, Journal of the American Medical Informatics Association, (2006)
- [2] A. Ragni et al. : Data Augmentation for Low Resource Languages, INTERSPEECH, (2014)
- [3] 大賀 隆裕, 矢野 正基, 大西 正輝 : 「Data and Test Time Augmentation によるアンサンブル学習の効果検証」, SIG Technical Reports, (2018)
- [4] 中村 寛 et al. : 「顧客要件に合致するパッケージシステムのノウハウ事例抽出支援システムの開発」, 人工知能学会第2種研究会資料, (2014)
- [5] 馬庭 伸栄 et al. : 「AI 技術を活用した SE 変革の実践」, FUJITSU. 68, (2017)
- [6] Janome v0.3 documentation (ja): <http://mocabeta.github.io/janome/>, (最終閲覧日 2018/11/07)
- [7] SAP TCodes : <https://www.sap-tcodes.org/>, (最終閲覧日 2018/11/07)
- [8] 経済産業省 スキルスタンダード本編 : http://www.meti.go.jp/policy/servicepolicy/contents/management_support/files/download.html, (最終閲覧日 2018/11/7)