

修士学位論文

題目

深層学習を利用した機能仕様書からの
ファンクションポイント計測支援

指導教員

楠本 真二 教授

報告者

小田郁弥

令和6年2月1日

大阪大学 大学院情報科学研究科

コンピュータサイエンス専攻

令和 5 年度 修士学位論文

深層学習を利用した機能仕様書からの
ファンクションポイント計測支援

小田郁弥

内容梗概

一般にシステム開発プロジェクトの見積もりでは、まず開発規模が見積もられ、それを用いて工数や予算の見積もりが行われる。開発規模の尺度として、ファンクションポイント（FP）が存在する。FP は客観的な見積もりの指標として広く用いられており、幾つかの計測方法は ISO/IEC の国際標準規格として扱われている。一方で、FP の計測は手作業で行われており、その計測の手間は FP の導入における課題となっており、その対策の一つとして FP 計測支援ツールに関する研究が行われている。本稿では、日本語で記述された機能仕様書を対象として、深層学習モデルを用いた FP 計測支援手法について提案する。提案手法は深層学習を用いた固有表現認識によって FP の構成要素を抽出・分類するものであり、その手法を実現する計測支援ツールを開発した。ツールの評価にあたっては、三種の題材を対象にツールを適用する実験を行った。一つ目の題材である FP 学習用の教材 2 種を用いて、それぞれ 19 要件と 16 要件から学習することで、それぞれ 4 要件中 1 件、5 要件中 1 件で FP の基本構成要素の分類結果が人間が分類した結果と同様になるという結果を得た。二つ目の題材であるデジタル庁の公開している 33 種の機能仕様書から抽出した要件を用いて 10 分割交叉実験を行い、平均して 215.4 要件中 192.4 要件で FP の基本構成要素の分類結果が人間が分類した結果と同様になるという結果を得た。三つ目の題材である総務省の公開している印鑑登録システムでは、機能仕様書から専門家が抽出した要件を用いて 10 分割交叉実験を行い、平均して 22 要件中 12.6 要件で FP の基本構成要素の分類結果が専門家が分類した結果と同様になるという結果を得た。

主な用語

自動見積もり、ファンクションポイント、深層学習、固有表現認識

目次

1	はじめに	1
2	準備	3
2.1	機能仕様書	3
2.2	ファンクションポイント (FP)	3
2.3	固有表現認識 (NER)	4
2.4	Bi-LSTM CRF	5
2.5	関連研究	6
3	提案手法	8
3.1	概要	8
3.2	提案手法	8
3.3	支援ツール	10
4	評価実験	13
4.1	実験目的	13
4.2	実験設定	13
4.3	評価方法	17
4.4	実験結果	19
5	考察	37
6	おわりに	38
	謝辞	39
	参考文献	40
	付録	43
7	小規模題材に用いた機能仕様書	44
7.1	作品管理システム	44
7.2	酒屋の在庫管理システム	44

図目次

1	FP 計測の流れ	5
2	NER による FP の抽出・分類	6
3	Bi-LSTM モデルのラベル割り当て	6
4	FP 計測におけるツール補助の概要	8
5	ツールを用いた処理の流れ	11
6	学習ツールの操作画面	12
7	タグ付けツールの操作画面	12
8	試験データの単語ごとのラベル一致数 (作品管理システム)	19
9	試験データの要件の個数一致数と完全一致数 (作品管理システム)	20
10	学習データの単語ごとのラベル一致数 (作品管理システム)	20
11	学習データの要件の個数一致数と完全一致数 (作品管理システム)	22
12	試験データの単語ごとのラベル一致数 (在庫管理システム)	22
13	試験データの要件の個数一致数と完全一致数 (在庫管理システム)	24
14	学習データの単語ごとのラベル一致数 (在庫管理システム)	24
15	学習データの要件の個数一致数と完全一致数 (在庫管理システム)	25
16	試験データの単語ごとのラベル一致数の平均 (題材 2)	27
17	試験データの要件の個数一致数と完全一致数の平均 (題材 2)	27
18	学習データの単語ごとのラベル一致数の平均 (題材 2)	28
19	学習データの要件の個数一致数と完全一致数の平均 (題材 2)	28
20	試験データの単語ごとのラベル一致数 (題材 3・SFP 法)	31
21	試験データの要件の個数一致数と完全一致数 (題材 3・SFP 法)	32
22	学習データの単語ごとのラベル一致数 (題材 3・SFP 法)	32
23	学習データの要件の個数一致数と完全一致数 (題材 3・SFP 法)	33
24	試験データの単語ごとのラベル一致数 (題材 3・NESMA 法)	34
25	試験データの要件の個数一致数と完全一致数 (題材 3・NESMA 法)	34
26	学習データの単語ごとのラベル一致数 (題材 3・NESMA 法)	36
27	学習データの要件の個数一致数と完全一致数 (題材 3・NESMA 法)	36

表目次

1	作成する学習データの例	9
2	形態素解析を行った学習データの例	10
3	題材の概要	13
4	作品管理システムに含まれる各ラベルの単語数	14
5	在庫管理システムに含まれる各ラベルの単語数	14
6	題材 2 に含まれる各ラベルの単語数	15
7	印鑑登録システムに含まれる各ラベルの単語数	16
8	個数一致となった要件の例	18
9	学習データと試験データの評価指標 (作品管理システム)	20
10	学習データの統計的指標 (作品管理システム)	21
11	試験データの統計的指標 (作品管理システム)	21
12	学習データと試験データの評価指標 (在庫管理システム)	22
13	学習データの統計的指標 (在庫管理システム)	23
14	試験データの統計的指標 (在庫管理システム)	23
15	学習データと試験データの評価指標の平均 (題材 2)	26
16	学習データの統計的指標の平均 (題材 2)	29
17	試験データの統計的指標の平均 (題材 2)	30
18	学習データと試験データの評価指標の平均 (題材 3・SFP 法)	32
19	学習データの統計的指標の平均 (題材 3・SFP 法)	33
20	試験データの統計的指標の平均 (題材 3・SFP 法)	33
21	学習データと試験データの評価指標の平均 (題材 3・NESMA 法)	34
22	学習データの統計的指標の平均 (題材 3・NESMA 法)	35
23	試験データの統計的指標の平均 (題材 3・NESMA 法)	35

1 はじめに

一般に、ソフトウェアシステム開発においては初期段階でのシステム開発規模の見積もりが重要である。不正確な見積もりはプロジェクト失敗の原因になることも報告されている。正確な見積もりの実現を目指し、システム開発規模の見積もりに関する研究が盛んに行われている [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12]。

システム開発における見積もりの指標として、ファンクションポイント (FP) [13] が存在する。FP は客観的な見積もりの指標として広く用いられており、IFPUG 法 [14] 等の幾つかの計測手法は ISO/IEC の国際標準規格 [15] となっている。通常、FP の計測は計測者が手動で行っている。そのためプロジェクトの規模に比例した計測コストが必要となることが、FP 導入の課題の一つとなっている [16] [17] [18]。文献 [19] によると、2004 年以降、FP 研究の焦点は FP の利点や評価から計測補助へと移行している。また、開発現場のニーズでも、汎用的な自動計測ツールや FP 教育手法の確立が求められている。従って、開発現場では FP 計測にかかるコストの低減と計測者を用意するために必要な教育コストが負担となっていると考えられる。また、自動計測ツールに関する研究も行われているが、特定組織の開発に依存しているものや、現場導入において手間のかかるものも多いという指摘もされている。

本研究では自然語（日本語）で記述されたシステムの機能仕様書に対し、深層学習モデルを用いた FP 計測支援手法を提案する。加えて、同手法を実現するツールを実装する。ツールの開発に当たっては、深層学習に関する専門知識がなくとも利用できる事を前提し、ユーザーが過去に作成した機能仕様書から深層学習モデルを作成できるようにした。提案手法では、ユーザの所有している機能仕様書を学習することで、語彙や文体などの組織のコンテキストを考慮した深層学習モデルを作成することで、組織のコンテキストに左右されない見積もりを行う。ツールのケーススタディとして、三種の題材を対象にツールを適用する実験を行う。一つ目の題材である FP 学習用の教材 2 種を用いて、それぞれ 19 要件と 16 要件から学習することで、それぞれ 4 要件中 1 件、5 要件中 1 件で FP の基本構成要素の分類結果が人間が分類した結果と同様になるという結果を得た。二つ目の題材であるデジタル庁の公開している 33 種の機能仕様書から抽出した要件を用いて 10 分割交叉実験を行い、平均 1938.6 件の要件から学習することで、平均して 215.4 要件中 192.4 要件で FP の基本構成要素の分類結果が人間が分類した結果と同様になるという結果を得た。三つ目の題材である総務省の公開している印鑑登録システムの機能仕様書では、専門家が抽出した要件を用いて 10 分割交叉実験を行い、平均 198 件の要件から学習することで、平均して 22 要件中 12.6 要件で FP の基本構成要素の分類結果が人間が分類した結果と同様になるという結果を得た。以降、2 章では研究の背景となる諸用語や関連研究について述べる。3 章では提案手法を用いた計測支援ツールについて述べる。4 章では実際にツールを用いて FP の基本構成要素の抽出・分類機能を確認するための実験とその結果について、5 章では考察について述べる。最後

に，6 章では本研究のまとめと今後の課題を述べる．

2 準備

本章では諸用語や関連研究について簡単に述べる。

2.1 機能仕様書

IEEE 標準 [20] では, requirements specification は, ソフトウェアシステムまたはコンポーネントの要求を規定した文書であり, 通常, 機能要求, 性能要求, インターフェイス要求, 設計要求, 開発標準などが含まれる。

requirements specification は日本語に訳すと, 要求仕様書, 要件仕様書, 要求定義書, 要件定義書等々, 色々な言い方があるが, ここでは, 日本情報システム・ユーザー協会に従い, 要求仕様書と呼ぶ [21]。内, 機能に関する仕様をまとめたものを本論文では機能仕様書と呼称する。

2.2 ファンクションポイント (FP)

ファンクションポイント (FP) はシステムの規模のメトリクスである [13]。FP 法はユーザーから見たシステムの各機能の量を計測する手法であり, システム開発における客観的な見積もりの指標として扱われている。FP 法は画面や帳票, ファイルなどを通じた情報の入出力に着目し, それらを種類別に数え上げ, 種類数を加重合計した値を機能量としている。FP 法は 1979 年に A. J. Albrecht によって提案されて以降, IFPUG 法 [14], NESMA 法 [22], SFP 法 [23] 等, 様々な FP 法が提案されている。以降, FP 法とは IFPUG 法を意味することとする。IFPUG 法はビジネスアプリケーションソフトウェアを対象に, 欧米で広く用いられている。IFPUG 法はシステムの持つ機能を意味のあるデータ集合であるデータファンクション (DF) と, 入出力や表示などのデータの流れであるトランザクションファンクション (TF) に分類する。DF は更に内部論理ファイルと外部論理ファイルの 2 種に, TF は外部入力, 外部出力, 外部照会の 3 種に分類される。これらの 5 つの分類は, FP の基本構成要素と呼称されている [24]。以下に詳細を示す。

1. 内部論理ファイル (ILF)

見積もり対象のシステム内部の書き換え可能なデータの集合

2. 外部論理ファイル (EIF)

見積もり対象のシステムによって参照されるが書き換えが行われないデータの集合

3. 外部入力 (EI)

データの書き込みや記録など, システム外部からの入力によって ILF の更新を伴う処理

4. 外部出力 (EO)

システム外部へのデータ出力のうち、出力に合計値や平均値の算出などのデータの加工を含むもの

5. 外部照会 (EQ)

システム外部へのデータ出力のうち、出力にデータの加工を必要とせず、ILF の更新処理を行わないもの

FP の点数の計測には、各基本構成要素の複雑度を計測し、各要点の点数を測定する。それらの点数の合計をシステムの機能の大きさとして評価する。図 2 に FP 計測の流れを示す。また、NESMA 法や SFP 法では、複雑さが考慮されず、基本構成要素の種類ごとに点数が固定されており、それらの個数から FP の点数を計測することができる。IFPUG 法では図 2 の FP 計測の際、各要件ごとに個別に、中、高のいずれかの複雑度を決定し、点数を算出する必要があるが、NESMA 法や SFP 法では、基本構成要素の種類のみで各要件毎の点数を決定することができる。機能仕様書には複雑度を評価するために十分な詳細が記載されていない場合があるため、開発初期の段階で IFPUG 法に準じて FP を計測する手段としてこれらの手法が存在する。

NESMA 法は NESMA (Netherlands Software Metrics Association) が提案した手法であり、FP 概算法 (estimated function point count) と FP 試算法 (indicative function point count) の 2 つの FP 計測法がある。FP 概算法は、先ず 全ての基本構成要素を抽出する。次に、全ての DF (ILF, EIF) の複雑度を低に、全ての TF (EI, EO, EQ) の複雑度を中とし、最後に FP を計算する。従って、IFPUG 法との違いは、複雑度の評価を基本構成要素毎に行うのではなく、上記のデフォルト値で決める点である。FP 試算法では、ILF と EIF の数を決定し、FP 値を $35 \times (\text{ILF の数}) + 15 \times (\text{EIF の数})$ で決定する即ち、FP 試算値は、IFPUG 法と異なり、論理ファイル (ILF, EIF) のみに基づいて算出される。

SFP 法は、2010 年に Roberto Meli によって提案された軽量な FP 計測手法である [25]。SFP は ISO14143-1 規格に準拠し IFPUG 法と互換性を持つ。SFP の計測では、計測対象から DF と TF の数を決定し、FP 値を $4.6 \times (\text{TF の数}) + 7 \times (\text{DF の数})$ で決定する。

NESMA 法、SFP 法と IFPUG 法の計測結果の比較も行われており、Meli らの調査 [25] では SFP と IFPUG の間で 0.9 以上の相関関係が確認されている。また、大岩らの調査 [26] では NESMA 概算法と IFPUG の間で 0.9 以上の相関関係が確認されている。本研究では、開発早期に作成される機能仕様書からの FP 計測手法として、NESMA 法、SFP 法を想定し、これらの手法による見積もりの支援を行う。

2.3 固有表現認識 (NER)

固有表現認識 (NER) は、1996 年の Message Understanding Conference で Grishman らが提唱した概念である [27]。NER は自然言語で記述されたテキスト内の固有の表現を識別し、事前に定義され

た組織名や人名といったラベルを付与する（カテゴリに分類する）自然言語処理のタスクであり，質疑応答 [28]，関係抽出 [29]，対話システム [30] など様々なタスクで応用されている．これにより，テキスト内の要素を自動的に抽出・分類することが可能となる．カテゴリの定義は目的に応じて自由に設定することができ，NER の手法によっては同じ単語でも前後の文脈から意味を判断することができる．例えば，"梅田駅に行くために十三人の客が十三駅で十三時に電車に乗った"のような文章があった場合，"人"の文脈を読み取ることで，最初の"十三"は数詞と判断できる．同様に"駅"の文脈から"梅田"と二つ目の"十三"は地名，"時"の文脈から三つ目の"十三"は時間と判断できる．

本研究では機能仕様書に記述された要件の文章のうち，図 2 に示すような形で FP の基本構成要素を抽出・分類するために使用する．

2.4 Bi-LSTM CRF

Bi-LSTM CRF (Bidirectional Long Short-Term Memory Conditional Random Fields) は，自然言語処理のタスクにおいて，系列ラベリング問題を解くモデルの一つである [31]．この手法は，Bidirectional Long Short-Term Memory (Bi-LSTM) と Conditional Random Fields (CRF) を組み合わせることにより，文脈を考慮した系列タグ付けを実現する．Bi-LSTM は，文脈の前後関係を考慮することができる再帰型ニューラルネットワーク (RNN) の一種である．このモデルは，図 3 に示すように前後の系列を考慮してラベルを割り当てることができるため，系列データ内の依存関係をモデリングするのに適している．CRF は，系列データの相互依存関係を考慮してラベルを割り当てる確率的グラフィカルモデルである．Bi-LSTM CRF は，Bi-LSTM で抽出された特徴量を基に，CRF を使用してラベルの系列に対する最適な推定を行う．固有認識表現の分野では広く用いられており，実例と

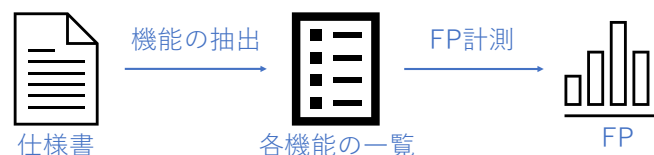


図 1 FP 計測の流れ

しては生物医学分野でのエンティティ分類を行った Hao らによる研究がある [32].

本実験では、上述の NER を行うための手法として本モデルを使用する。

2.5 関連研究

自然語で記述された機能仕様書からの FP 計測を対象とした関連研究について説明する。

KUI ZHANG ら [33] は深層学習モデルを利用することで、中国語で記述された機能仕様書内の要件からファンクションポイントの基本構成要素を抽出・分類し、システムの規模の推定を支援する研究を行った。同研究では、29 の実プロジェクトの機能仕様書を用いた実験を行い、深層学習モデルは

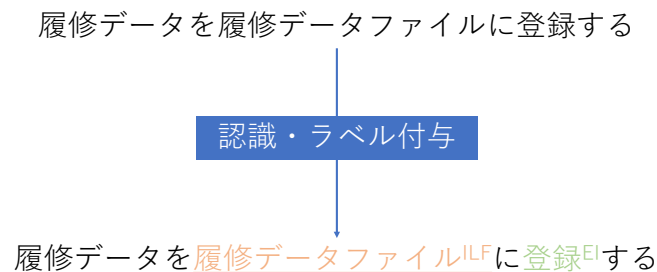


図2 NER による FP の抽出・分類

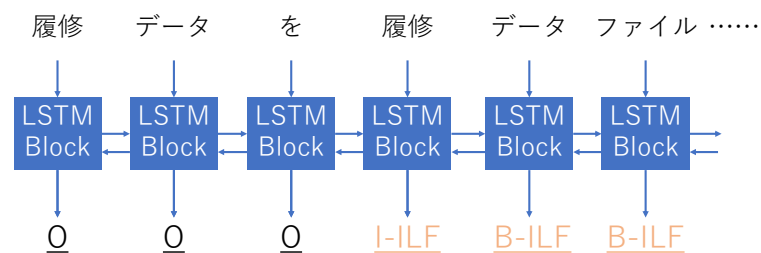


図3 Bi-LSTM モデルのラベル割り当て

94.25% の精度で FP の基本構成要素を抽出・分類することができ、人力での基本構成要素の分類にかかる時間を大きく削減できる可能性があると報告している。但し、この研究に用いられた機能仕様書は公開されておらず、ツールの公開も行われていない。

また、文献 [34] では、日本語の構文解析 (係り受け解析) ツールの cabocha を用いて、日本語で記述された機能仕様書に対するファンクションポイントの計測支援を行っている。機能仕様書から機能仕様についての文章を提案されているテンプレートに沿って書き換え、その仕様を自動解析ツールを用いて解析し計測に必要な要素を抽出している。しかし、テンプレートへの書き換え部分の手間が大きいという課題がある。

3 提案手法

3.1 概要

FP の計測は計測者が手動で行っている．そのためプロジェクトの規模に比例した計測コストが必要となることから，FP 導入の課題の一つとなっている．文献 [19] によると，2004 年以降，FP 研究の焦点は FP の利点や評価から計測補助へと移行している．また，開発現場のニーズでも，汎用的な自動計測ツールや FP 教育手法の確立が求められている．従って，開発現場では FP 計測にかかるコストの低減と計測者を用意するために必要な教育コストが負担となっていると考えられる．また，自動計測ツールに関する研究も行われているが，特定組織の開発に依存しているものや，現場導入において手間のかかるものも多いという指摘もされている．

本研究では日本語で記述されたシステムの機能仕様書に対し，深層学習モデルを用いてファンクションポイントの計測支援を行う手法を提案する．これにより，手動で行っている FP 計測作業の一部ないし全体を自動化することができ，計測コストと計測者の負担の低減が可能である．

3.2 提案手法

提案手法はモデルの作成と，FP の基本構成要素の抽出・分類の二つの段階に大別される．まず，学習データとなる機能仕様書の各単語にラベルを付与した分類済機能仕様書を用いて学習モデルを作成する．次に，作成した学習モデルと FP 計測の対象となる要件が記入された機能仕様書にラベルが付与して要件を抽出・分類する．以下に詳細を示す．

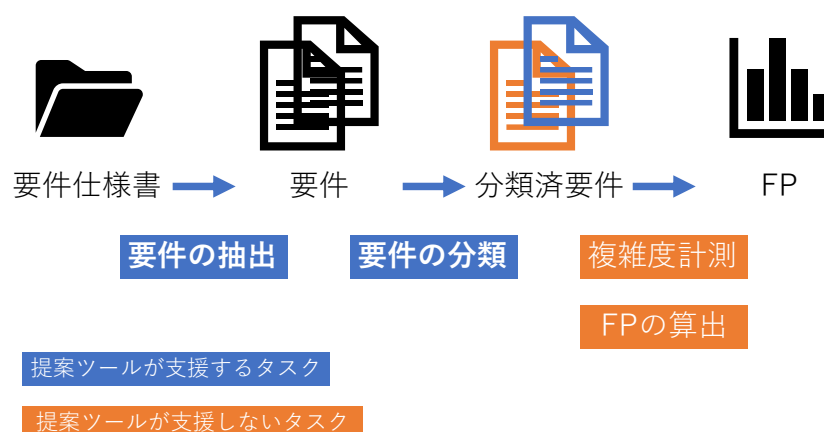


図 4 FP 計測におけるツール補助の概要

3.2.1 モデルの作成

学習データを基に Bi-LSTM CRF モデルを作成する。学習データは要件の文章に FP の基本構成要素のラベルを付与したものが与えられる。更に、それらの文章に対して形態素解析を行い、単語レベルに分割する。モデルの学習にはこちらの形態素解析を行ったデータを用いる。学習データの例を表 1 に、形態素解析を行い、単語レベルに分割した例を表 2 に記す。

ラベル NESMA 法による計測では 5 つの基本構成要素 (EI, EO, EQ, ILF, EIF) に対して、文頭を意味する I を付与したもの (例えば, I-ILF) と、前の単語から継続する事を表す B を付与したもの (例えば, B-ILF) の計 10 種類に加え、基本構成要素に属さない事を示す O の 11 種類のラベルが存在する。表 1 の文章では、"データベース"が内部論理ファイルであるため ILF に、"保存"は ILF のデータを更新する動作なので EI に分類される。このようにデータを用意すると、ツールが形態素解析を行う。例えば"データベース"は"データ"と"ベース"に分類され、文頭の"データ"には ILF に I を付与した I-ILF が、"ベース"には前の単語から継続する事を表す B を付与した B-ILF のタグが付与される。加えて、必要に応じてラベルを追加することで、FP の基本構成要素以外の情報の抽出も可能となる。

表 1 と表 2 の例では、"データベース"が ILF、"保存"が EI にあたる。形態素解析を行うことで、"データベース"は"データ"と"ベース"の 2 語に分割される。このとき、ILF の付与された文章の文頭に当たる"データ"の側には I-ILF のタグが付与され、そのあとに続く語である"ベース"には B-ILF のタグが付与される。"保存"は一語で EI を表しているが、このような場合は I-EI が付与される。

また、必要に際して他のラベルを追加することで、ユーザーにとって必要な情報の抽出にも用いることができるほか、基本構成要素のラベルを別のものに置き換えてモデルを作成することもできる。例えば、シンプルファンクションポイント (SFP) の計測においては、DF と TF の二つで十分である。この際はラベルに基本構成要素ではなく、DF と TF、それ以外を示す O のラベルを用いることで学習データを作成することができる。

表 1 作成する学習データの例

文章	ラベル
入力されたテキストを	O
データベース	ILF
に	O
保存	EI
する	O

3.2.2 FP の基本構成要素の抽出・分類

前項で作成したモデルに、機能仕様書の要件の文章を入力として与える。入力された文章を形態素解析により単語ごとに分割し、各単語にラベルを付与することで基本構成要素の抽出・分類を行う。

ラベルの個数を計測することで、NESMA 法や SFP 法による FP の点数の算出が可能になる。また、複雑度の算出を行う際にも、FP の基本構成要素の対象となる単語が明らかになっていることから、見積もり作業のコストの低減が期待できる。

3.3 支援ツール

手法を実現するツール、FACTUAL を実装した。このツールはユーザーの端末上で Bi-LSTM CRF モデルを作成し、ファンクションポイントの基本構成要素を抽出・分類する。図 4 に IFPUG 法での FP 計測の流れと、FACTUAL が補助するタスクについて述べる。現在、FACTUAL が補助するのは機能仕様書内の機能要件の抽出と、抽出した機能要件の分類である。本ツールはユーザーのローカル環境内に導入することができ、学習に際してユーザーの過去に作成した機能仕様書を作成するため、組織の開発に依存せず、簡便に導入することが可能である。動作環境は Window11 で正常に動作することを確認している。

表 2 形態素解析を行った学習データの例

単語	ラベル
入力	O
された	O
テキスト	O
を	O
データ	I-ILF
ベース	B-ILF
に	O
保存	I-EI
する	O

3.3.1 ツールの概要

FACTUAL は Python で実装された FP 計測支援ツールであり、GitHub 上で公開されている*1。先述した提案手法を実装するために、深層学習部分では torch を、形態素解析にはオープンソース 形態素解析エンジンである MeCab を用いた。また、直感的に操作するために tkinter を用いたウィンドウアプリケーションとして本ツールを実装した。

FACTUAL は提案手法を実現するための二つのツールを内包している。図 5 にツールの利用の流れを示す。まず、学習データとなる機能仕様書の各単語にラベルを付与した分類済機能仕様書を学習ツールに与えることで学習モデルを作る。次に、作成した学習モデルと FP 計測の対象となる要件が記入された機能仕様書をタグ付けツールに入力することで、ラベルが付与され要件が出力される。以下に詳細を示す。

3.3.2 学習ツール

学習ツールはモデルの作成を行うツールである。ユーザーに表示される画面を図 6 に示す。提案手法を実現するために、CSV 形式、CONLL 形式のファイルを読み込み、torch を用いて深層学習モデルの作成を行う。表 1 の形の CSV データを入力することで、自動的に形態素解析が行われ、表 2 の形の CONLL ファイルが作成される。一度作成した CONLL ファイルは再度学習を行う際に再利用することが可能である。学習が完了すると、タグ付けツールで利用するモデルとデータファイルが出力される。

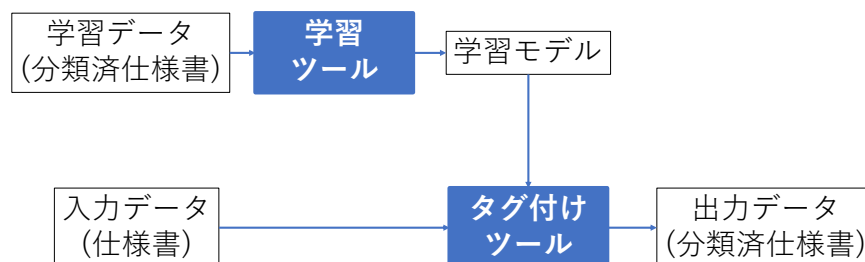


図 5 ツールを用いた処理の流れ

*1 <https://github.com/OdaFumiya/FACTUAL>

モデルの作成は、ユーザーが保有している機能仕様書などの過去の開発資料を利用することを想定している。これにより、ユーザーの文章の書式や使用する語彙の傾向をモデルに反映することができるため、推測の精度向上を見込むことができる。

3.3.3 タグ付けツール

タグ付けツールは FP の基本構成要素の抽出・分類を行うツールである。ユーザーに表示される画面を図 7 に示す。学習ツールで作成したモデルとデータファイル、そして要件が記述された機能仕様書を入力することでラベル付された単語が CSV 形式で出力される。入力には TXT 形式のファイルが対応している。

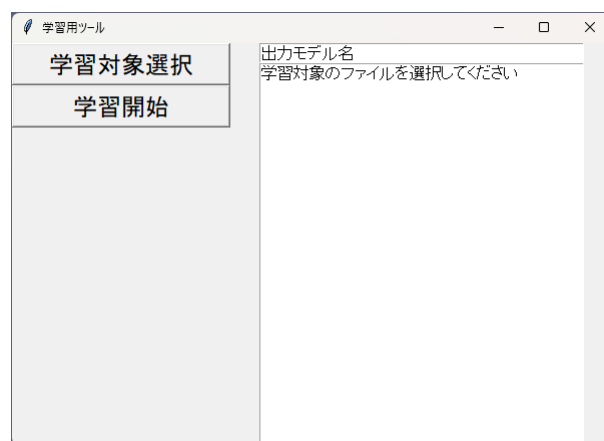


図 6 学習ツールの操作画面

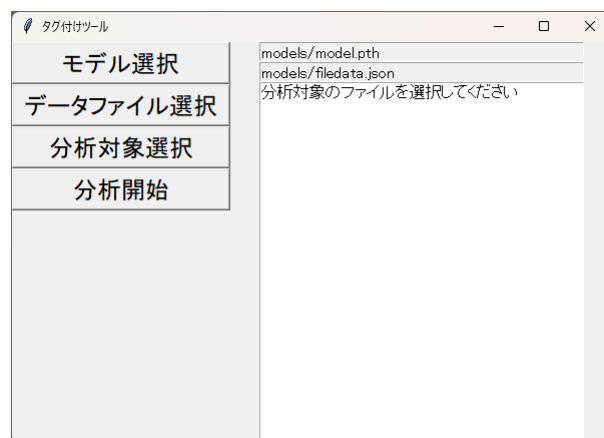


図 7 タグ付けツールの操作画面

4 評価実験

4.1 実験目的

提案手法の FP の基本構成要素の抽出・分類機能の精度を調査するために、後述する題材に提案ツールを用いて実験を行った。本実験では学習によって FP の基本構成要素の抽出・分類が行えることの確認を目的とし、学習の経過と、学習の結果の評価を示す。

以下に実験の詳細を示す。

4.2 実験設定

提案ツールを用いて、以下の題材から作成した学習データを学習してモデルを作成し、作成したモデルで試験データにラベル付を行い、後述の評価指標を計測する。学習の進行度ではエポックと呼ばれる単位を用いる。学習は学習データを読み込み、モデルのパラメーターの更新を行うことで行われる。学習データ全てを読み込み、パラメーターを更新することで 1 エポックの学習が完了する。

実験題材として以下の三つの題材を用意した。表 3 に概要を示し、以降に詳細を示す。

4.2.1 題材 1

FP の学習用に作成された二つの機能仕様書を用いて実験を行った。機能仕様書は DVD 店の作品管理システムと酒屋の在庫管理システムを題材とした小規模なものであり、それぞれ 23 要件と 21 要件が含まれる。機能仕様書の内容については付録である 7 章に掲載する。

学習データと試験データに含まれる各ラベルの単語数を表 4 と表 5 に記す。前者は 19 要件を学習データ、4 要件を試験データとし、後者は 16 要件を学習データ、5 要件を試験データとし、200 エポックの学習を行う。5 種類の FP の基本構成要素について後述の評価指標を計測する。学習の際のモデルのバッチサイズは 4、隠れ層は 32、学習レートは 0.01 とし、最適化アルゴリズムは Adam を用いた。

表 3 題材の概要

	概要	計測者	要件数	単語数
題材 1	FP の学習用に作成された 2 つの機能仕様書	筆者	23+21 件	368+641 語
題材 2	デジタル庁の提供する 33 の機能仕様書	筆者	計 2154 件	計 51830 語
題材 3	総務省の提供する 1 つの機能仕様書	CFPS Fellow	220 件	8484 語

表 4 作品管理システムに含まれる各ラベルの単語数

ラベル	学習データ	試験データ
I-ILF	15	3
B-ILF	23	3
I-EIF	0	0
B-EIF	0	0
I-EI	8	1
B-EI	10	2
I-EO	5	2
B-EO	5	2
I-EQ	4	1
B-EQ	1	1
O	235	47
合計	306	62

表 5 在庫管理システムに含まれる各ラベルの単語数

ラベル	学習データ	試験データ
I-ILF	15	6
B-ILF	31	13
I-EIF	16	9
B-EIF	19	5
I-EI	4	1
B-EI	4	1
I-EO	4	2
B-EO	4	2
I-EQ	1	1
B-EQ	3	1
O	394	102
合計	495	146

4.2.2 題材 2

デジタル庁の提供するデータ要件・連携要件の標準仕様^{*2}に含まれる 33 プロジェクト 2154 要件を用いて実験を行った。この機能仕様書の各要件に表 1 の形で筆者が手動でラベル付を行い、実験用のデータを作成した。要件全体の 90% を学習データ、残りを試験データとして 150 エポックの学習を行う 10 分割交差検証を行い、各指標の平均を評価する。

学習データと試験データに含まれる各ラベルの単語数を表 6 に記す。実験に用いたデータはシステム間の連携に関する機能仕様書であったため、5 種類の FP の基本構成要素にに加え、調査対象のシステムであることを意味するラベルの TS と、連携先のシステムである事を意味するラベルの ES を追加し、調査を行った。これにより、ユーザーが FP の基本構成要素以外の情報の抽出を行うことができるかどうかの確認を行うことができる。学習の際のモデルのバッチサイズは 4、隠れ層は 32、学習レートは 0.001 とし、最適化アルゴリズムは Adam を用いた。

表 6 題材 2 に含まれる各ラベルの単語数

ラベル	項目数
I-TS	2151
B-TS	6870
I-ES	2180
B-ES	6320
I-ILF	550
B-ILF	2424
I-EIF	1643
B-EIF	7229
I-EI	291
B-EI	0
I-EO	786
B-EO	40
I-EQ	1077
B-EQ	35
O	20233
合計	51830

^{*2} https://www.digital.go.jp/policies/local_governments/specification/

4.2.3 題材 3

FP を適切に計測できる技術を有することを認定する資格である CFPS を、20 年以上保有することで認定される CFPS Fellow [35] の有資格者が、総務省の提供する印鑑登録システム標準仕様書【第 3.1 版】*3から抽出した 220 要件を用いた。抽出された要件から実験用データを作成し、要件全体の 90% を学習データ、残りを試験データとして 200 エポックの学習を行う 10 分割交差検証を行い、各指標の平均を評価する。データに含まれる各ラベルの単語数を表 7 に記す。5 種類の FP の基本構成要素に加え、DF 内の一項目であることを示すラベルである item を追加し、調査を行った。また、5 種類の FP 基本構成要素による評価だけでなく、SFP で用いられる TF と DF の 2 種類の基本構成要素による評価も行った。TF の項目数は EI, EO, EQ の項目数の合計であり、DF の項目数は ILF と EIF の項目数の合計である。学習の際のモデルのバッチサイズは 4、隠れ層は 32、学習レートは 0.01 とし、最適化アルゴリズムは Adam を用いた。

表 7 印鑑登録システムに含まれる各ラベルの単語数

ラベル	項目数
I-item	122
B-item	239
I-ILF	25
B-ILF	27
I-EIF	0
B-EIF	0
I-EI	47
B-EI	297
I-EO	2
B-EO	16
I-EQ	35
B-EQ	209
O	7463
合計	8484

*3 https://www.soumu.go.jp/main_content/000900552.pdf

4.3 評価方法

本実験では以下の 4 つの指標を基に評価する。

4.3.1 各ラベルごとの統計的指標

各単語のラベルの予測の正答率を、学習データ、試験データそれぞれについて、各ラベル毎に F1 値、適合率、再現率を算出することで評価する。算出はそれぞれの題材で規定回数の学習が終了した時点で行う。いずれの指標も 0 以上 1 以下の値であり、精度が高ければ大きな値を取る。

4.3.2 単語ごとのラベル一致数

文章内の各単語に対して、設定したラベルと予測したラベルが一致する単語数を各エポックごとに評価する。正確な予測ができていればラベルの一致数が増加する。

4.3.3 要件ごとのラベルの完全一致数

各要件のうち、含まれる単語全ての予測したラベルが設定したラベルと一致している要件の個数を各エポックごとに評価する。正確な予想ができていればラベルの一致数が増加する。

4.3.4 要件ごとのラベルの個数一致数

各要件のうち、要件内の予測した基本構成要素のラベルが設定した基本構成要素のラベルと一致しているかを各エポックごとに評価する。具体的には両者の各ラベルの先頭要素である "I- (ラベル名) " の個数を比較する。これは NESMA 法や SFP 法の見積もりでは基本構成要素の個数が重要であること、また、ラベルの予測の際に見積もりには影響しないような軽微な付与の揺れを許容するためにこれを調査する。

完全一致ではないが個数一致である要件について、表 8 に記す。この例のように、修飾する語を含んでラベルを付与するか否かなどの軽微な揺れをこの評価指標で吸収することができる。

表 8 個数一致となった要件の例

単語	正解ラベル	予測ラベル
国民	I-TS	I-TS
健康	B-TS	B-TS
保険	B-TS	B-TS
システム	B-TS	B-TS
が	O	O
次期	I-ES	I-ES
国保	B-ES	B-ES
総合	B-ES	B-ES
システム	B-ES	B-ES
に	O	O
給付	I-EIF	O
管理	B-EIF	O
に	B-EIF	O
必要な	B-EIF	O
各種	B-EIF	I-EIF
情報	B-EIF	B-EIF
を	O	O
照会	I-EQ	I-EQ
する	O	O

4.4 実験結果

以下に、各題材での実験の結果を示す。

4.4.1 題材 1 に対する結果

作品管理システムに対する実験結果のグラフを図 8～11 に、在庫管理システムに対する実験結果のグラフを図 12～15 に示す。また、200 エポックの学習終了時点での作品管理システムに対する各評価指標の数値と各ラベル毎の統計的指標を表 9、表 10、表 11、在庫管理システムに対する結果を表 12、表 13、表 14 に記す。いずれの場合も学習データの学習はすべての単語についかなり早い段階で学習ができており、学習によって全ての基本構成要素の抽出・分類が行えている。しかし、試験データでは単語のラベル一致数は高いものの、要件の評価はあまり高くない事が見て取れる。両題材とも、単語のラベル一致数に対し、要件の一致数は低くなっている。これは要件の一致数判定に関わらない O タグが文章の大部分を占めており、重要な構成機能要素の単語数が少なく、うまく学習ができていないためと考えられる。特に EQ についてはどちらの題材にも出現数が少なく、学習ができていないことがわかる。KUI ZHANG らの先行研究 [33] と比較して大きく精度が劣ることから、学習データの不足が顕著に影響を及ぼしていると考えられる。

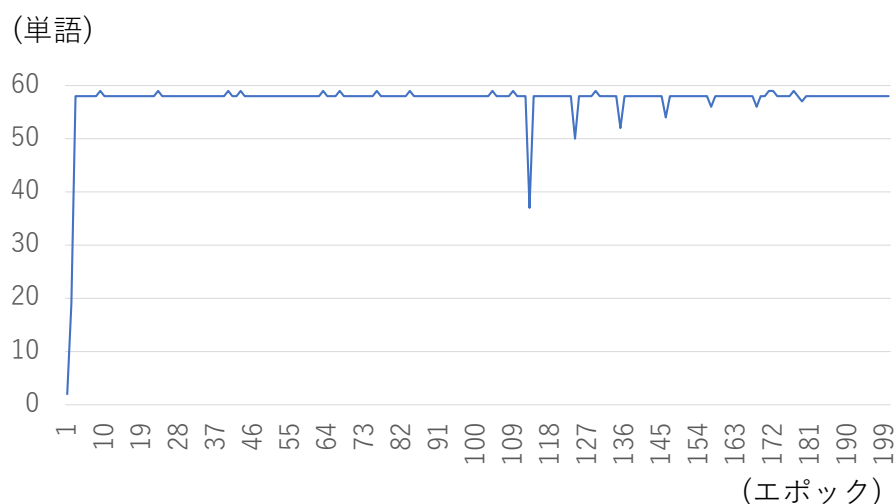


図 8 試験データの単語ごとのラベル一致数 (作品管理システム)

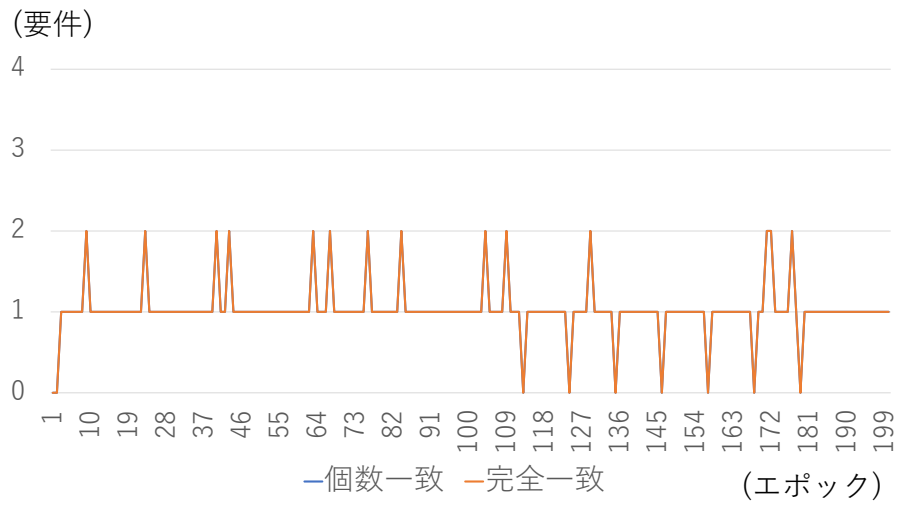


図9 試験データの要件の個数一致数と完全一致数 (作品管理システム)

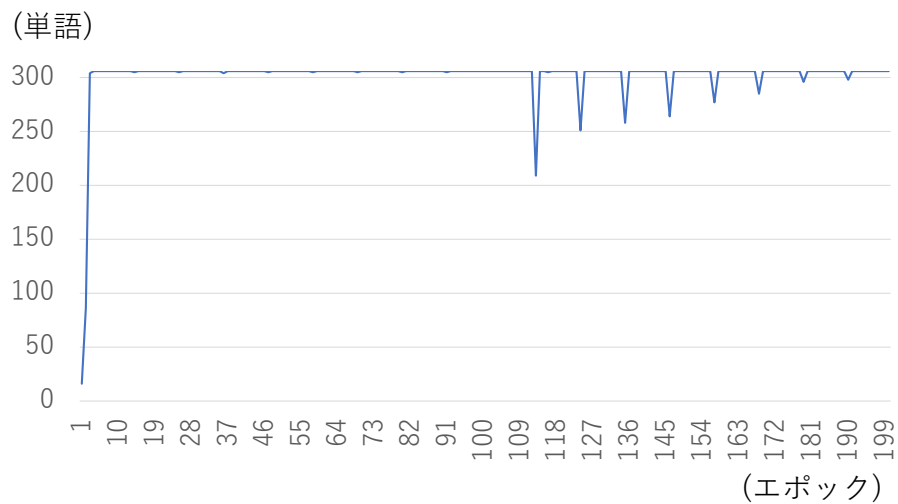


図10 学習データの単語ごとのラベル一致数 (作品管理システム)

表9 学習データと試験データの評価指標 (作品管理システム)

評価指標	学習データ	試験データ
単語ごとの一致数	306/306	58/62
要件ごとの完全一致数	19/19	1/4
要件ごとの個数一致数	19/19	1/4

表 10 学習データの統計的指標 (作品管理システム)

ラベル	F1 値	再現率	適合率
I-ILF	1.0000	1.0000	1.0000
B-ILF	1.0000	1.0000	1.0000
I-EIF	1.0000	1.0000	1.0000
B-EIF	1.0000	1.0000	1.0000
I-EI	1.0000	1.0000	1.0000
B-EI	1.0000	1.0000	1.0000
I-EO	1.0000	1.0000	1.0000
B-EO	1.0000	1.0000	1.0000
I-EQ	1.0000	1.0000	1.0000
B-EQ	1.0000	1.0000	1.0000
O	1.0000	1.0000	1.0000

表 11 試験データの統計的指標 (作品管理システム)

ラベル	F1 値	再現率	適合率
I-ILF	0.8000	0.6666	1.0000
B-ILF	1.0000	1.0000	1.0000
I-EIF	0.0000	0.0000	0.0000
B-EIF	0.0000	0.0000	0.0000
I-EI	0.0000	0.0000	0.0000
B-EI	1.0000	1.0000	1.0000
I-EO	0.8000	1.0000	0.6666
B-EO	0.8000	1.0000	0.6666
I-EQ	0.0000	0.0000	0.0000
B-EQ	0.0000	0.0000	0.0000
O	0.9791	1.0000	0.9591

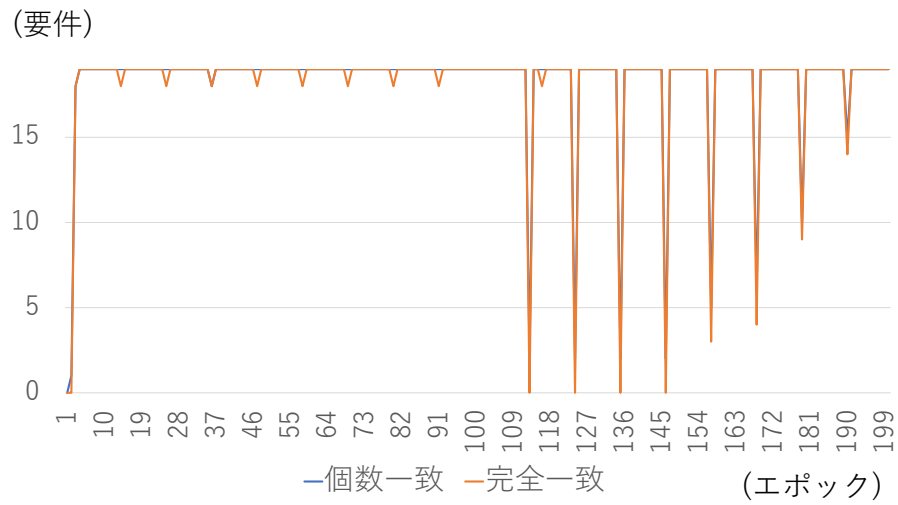


図 11 学習データの要件の個数一致数と完全一致数 (作品管理システム)

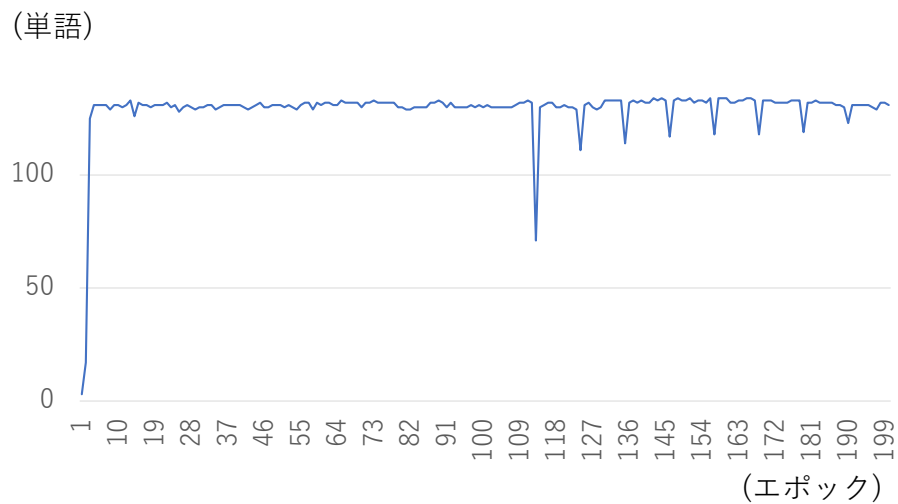


図 12 試験データの単語ごとのラベル一致数 (在庫管理システム)

表 12 学習データと試験データの評価指標 (在庫管理システム)

評価指標	学習データ	試験データ
単語ごとの一致数	495/495	131/146
要件ごとの完全一致数	16/16	0/5
要件ごとの個数一致数	16/16	1/5

表 13 学習データの統計的指標 (在庫管理システム)

ラベル	F1 値	再現率	適合率
I-ILF	1.0000	1.0000	1.0000
B-ILF	1.0000	1.0000	1.0000
I-EIF	1.0000	1.0000	1.0000
B-EIF	1.0000	1.0000	1.0000
I-EI	1.0000	1.0000	1.0000
B-EI	1.0000	1.0000	1.0000
I-EO	1.0000	1.0000	1.0000
B-EO	1.0000	1.0000	1.0000
I-EQ	1.0000	1.0000	1.0000
B-EQ	1.0000	1.0000	1.0000
O	1.0000	1.0000	1.0000

表 14 試験データの統計的指標 (在庫管理システム)

ラベル	F1 値	再現率	適合率
I-ILF	0.8333	0.8333	0.8333
B-ILF	0.8888	0.9230	0.8571
I-EIF	0.4285	0.3333	0.6000
B-EIF	0.5000	0.4000	0.6666
I-EI	1.0000	1.0000	1.0000
B-EI	1.0000	1.0000	1.0000
I-EO	1.0000	1.0000	1.0000
B-EO	0.6666	0.5000	1.0000
I-EQ	0.0000	0.0000	0.0000
B-EQ	0.0000	0.0000	0.0000
O	0.9528	0.9901	0.9181

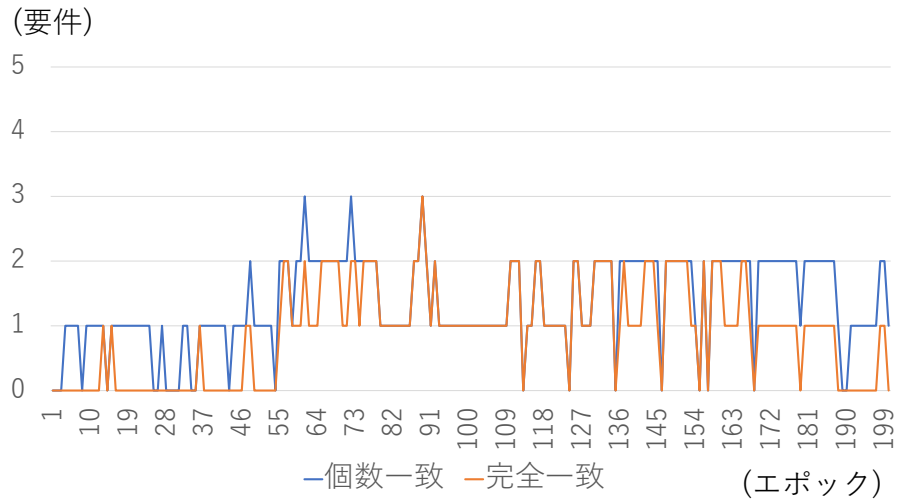


図 13 試験データの要件の個数一致数と完全一致数 (在庫管理システム)

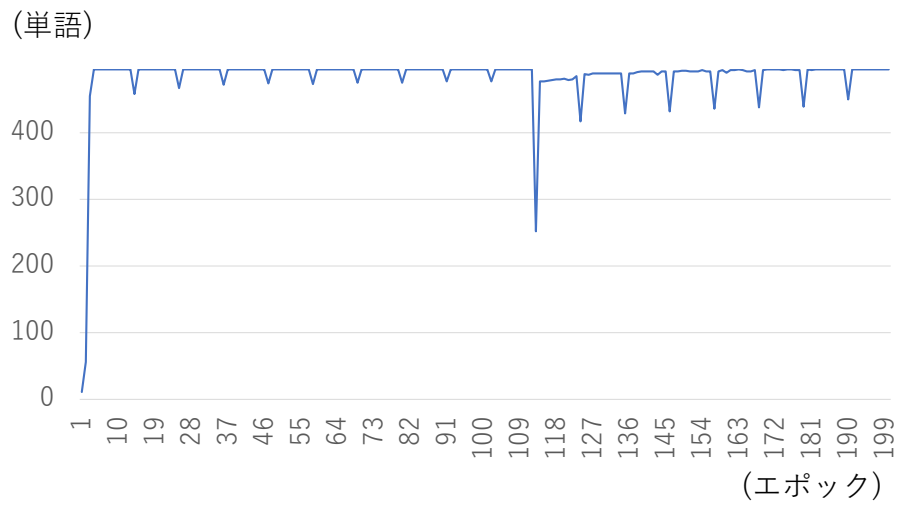


図 14 学習データの単語ごとのラベル一致数 (在庫管理システム)

(要件)

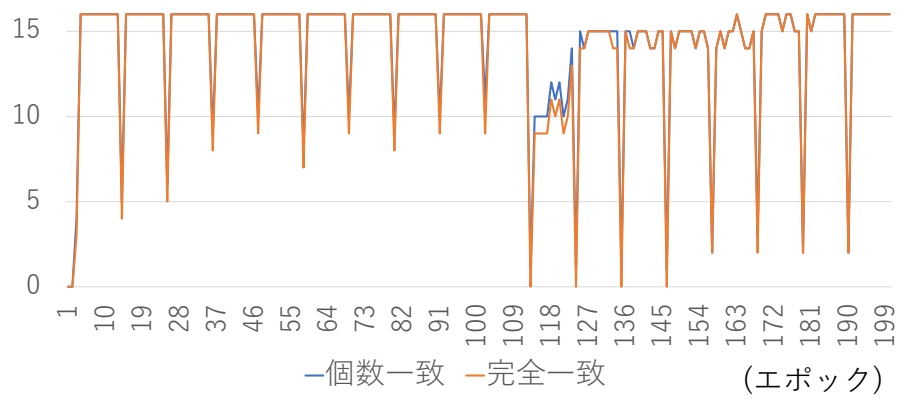


図 15 学習データの要件の個数一致数と完全一致数 (在庫管理システム)

4.4.2 題材 2 に対する結果

実験結果のグラフを図 16～19 に示す。いずれも 60 エポック程度までは大きな変動はないが、その後、徐々に右肩上がりに精度が向上していることが見て取れる。このことから、提案手法によって機能仕様書を学習し、要件の抽出・分類が可能であると考えられる。しかし、周期的に精度が大きく低下している点や、最初の 60 エポックまで精度に変動がないことなど、現状説明が不能な点がいくつか存在しており、そのことについては更なる調査が必要と考える。

150 エポックの学習終了時点での各評価指標の数値と各ラベル毎の統計的指標を表 15、表 16、表 17 に記す。各指標について、学習データでは約 97% の要件に対して、完全に正しくラベルを付与できていることがわかる。試験データでは約 85% の要件について完全に正しくラベルを付与することができしており、約 89% の要件に対してラベルの個数が正しいという結果になった。学習データの各ラベルの F1 スコアは全て 0.9 を上回っており、正しく学習できていることが読み取れる。試験データに関しては B-EO と B-EQ のラベルの F1 スコアが大幅に低くなっている。これは B-EO が含まれる単語が 40 語、B-EQ が含まれる単語が 35 語しか存在しないという、データの不足によるものが原因と考えられる。

また、試験データの F1 スコアのうち、DF である ILF と EIF の数値が他と比べてやや低くなっている。これに関しては文章の前後関係を踏まえて学習を行う Bi-LSTM CRF の影響と考えられる。今回の実験に用いた文章の多くは、"対象システムが外部システムからデータを受信する"のような形で記述されている。そのため、TS や ES のタグは順番が固定化されており、前後関係を基に判断がしやすい。TF にあたる EI, EO, EQ はそれぞれ"入力", "出力", "照会"等、各ラベルごとの語彙が少なく、また、他のラベルと競合することがないため、高い精度で分類できると考えられる。他方、DF は同じ位置に存在し、同じような語彙が異なるラベルに混在することから、ILF と EIF の混同が見受けられた。その結果、F1 スコアが他のラベルよりも低くなったと考えられる。

表 15 学習データと試験データの評価指標の平均 (題材 2)

評価指標	学習データ	試験データ
単語ごとの一致数	46444.5/46647.0	5031.0/5183.0
要件ごとの完全一致数	1872.9/1938.6	182.7/215.4
要件ごとの個数一致数	1907.1/1938.6	192.4/215.4

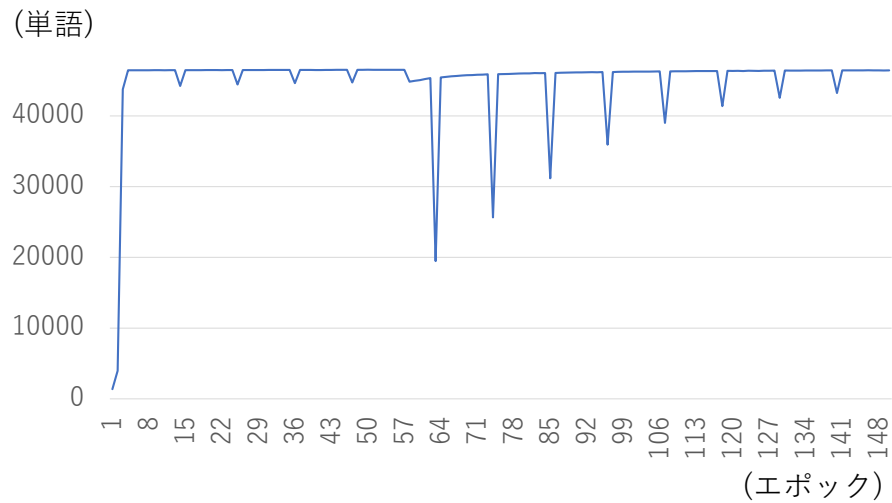


図 16 試験データの単語ごとのラベル一致数の平均 (題材 2)

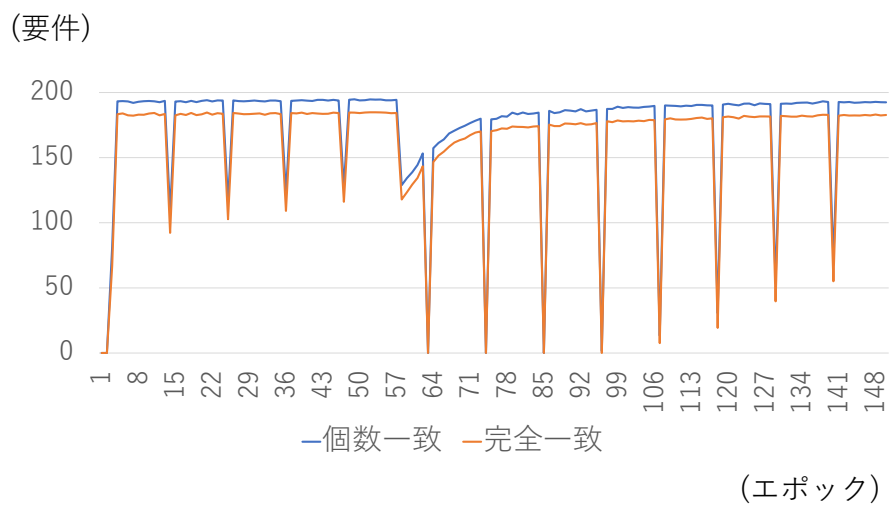


図 17 試験データの要件の個数一致数と完全一致数の平均 (題材 2)

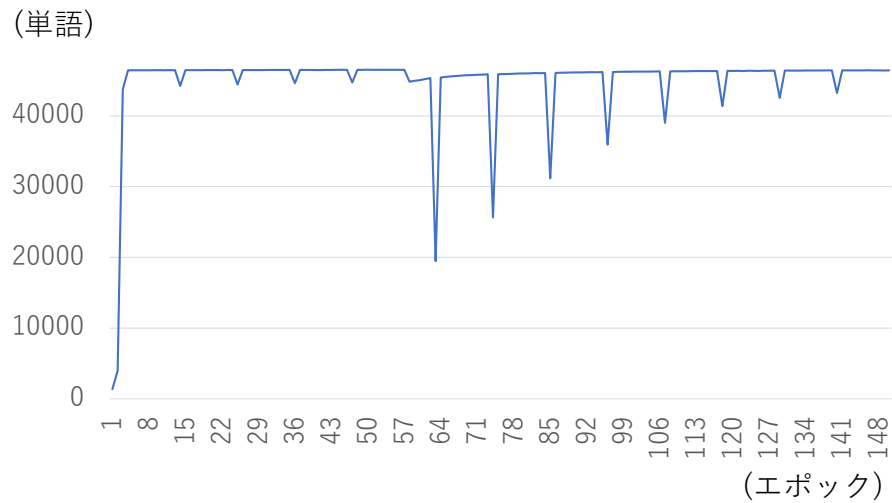


図 18 学習データの単語ごとのラベル一致数の平均 (題材 2)

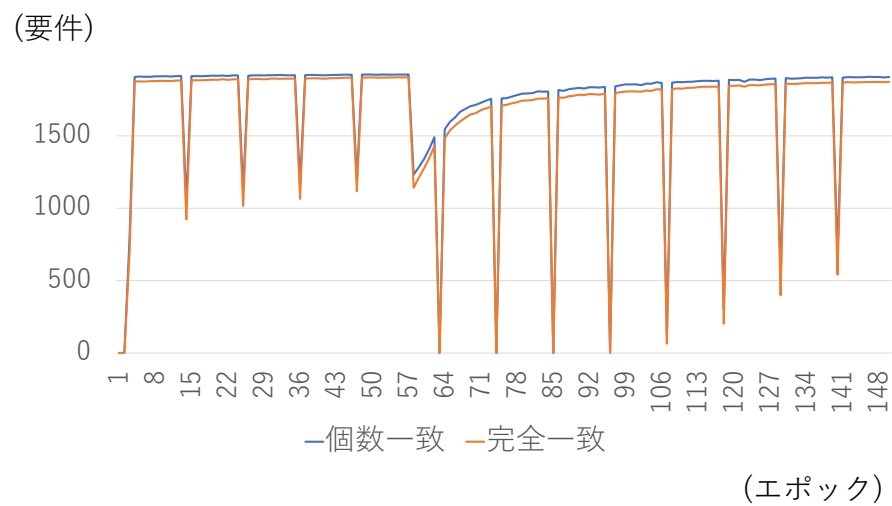


図 19 学習データの要件の個数一致数と完全一致数の平均 (題材 2)

表 16 学習データの統計的指標の平均 (題材 2)

ラベル	F1 値	再現率	適合率
I-TS	0.9988	0.9994	0.9981
B-TS	0.9992	0.9996	0.9987
I-ES	0.9973	0.9979	0.9966
B-ES	0.9978	0.9984	0.9972
I-ILF	0.9759	0.9808	0.9710
B-ILF	0.9843	0.9941	0.9746
I-EIF	0.9870	0.9844	0.9896
B-EIF	0.9932	0.9906	0.9958
I-EI	0.9935	0.9874	0.9996
B-EI	0.0000	0.0000	0.0000
I-EO	0.9965	0.9997	0.9933
B-EO	0.7918	0.6699	0.9821
I-EQ	1.0000	1.0000	1.0000
B-EQ	0.9878	0.9760	1.0000
O	0.9969	0.9966	0.9973

表 17 試験データの統計的指標の平均 (題材 2)

ラベル	F1 値	再現率	適合率
I-TS	0.9972	0.9986	0.9958
B-TS	0.9979	0.9993	0.9965
I-ES	0.9874	0.9882	0.9868
B-ES	0.9884	0.9879	0.9890
I-ILF	0.8354	0.8130	0.8606
B-ILF	0.8559	0.8471	0.8723
I-EIF	0.9133	0.9169	0.9098
B-EIF	0.9384	0.9395	0.9380
I-EI	0.9933	0.9869	1.0000
B-EI	0.0000	0.0000	0.0000
I-EO	0.9943	0.9988	0.9899
B-EO	0.4870	0.4661	0.5600
I-EQ	0.9995	0.9991	1.0000
B-EQ	0.4706	0.4545	0.5000
O	0.9823	0.9836	0.9811

4.4.3 題材 3 に対する結果

SFP 法に則って作成したデータをもとに行った実験結果のグラフを図 20～23 に、NESMA 法に則って作成したデータをもとに行った実験結果のグラフを図 24～23 に示す。また、200 エポックの学習終了時点での SFP 法で実験を行った各評価指標の数値と各ラベル毎の F1 スコアを表 18、表 19、表 20 に NESMA 法で実験を行った各評価指標の数値と各ラベル毎の F1 スコアを表 21、表 22、表 23 に記す。小規模題材に対する実験と同様に、いずれの場合も学習データの学習はすべての単語についかなり早い段階で学習ができており、学習によって 90% 以上の基本構成要素の抽出・分類が行えており、試験データでは単語のラベル一致数は高いものの、要件の評価はあまり高くない事が見て取れる。

試験データでは約 83% の要件について完全に正しくラベルを付与することができており、約 50% の要件に対してラベルの個数が正しいという結果になった。学習データの要件が存在する各ラベルの F1 スコアは全て 0.9 を上回っており、正しく学習できていることが読み取れる。試験データに関しては NESMA, SFP 双方で item 以外のラベルの F1 スコアが低くなっている。これは文脈によらずに単語によってラベルを特定できる item ラベルと、文脈から判断する必要のある他のラベルの違いであると考えられる。

また、SFP 法による計測と NESMA 法による計測では、要件の抽出に関してはやや SFP 法の方が精度が高いという結果が得られた。

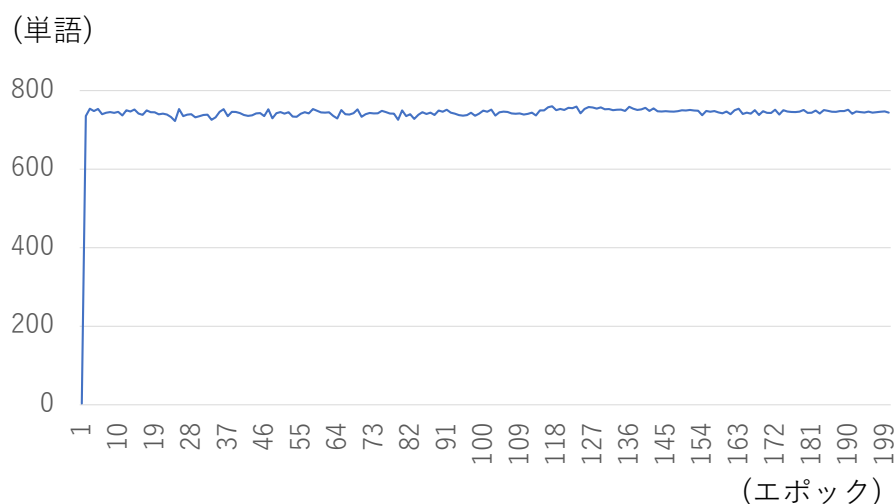


図 20 試験データの単語ごとのラベル一致数 (題材 3・SFP 法)

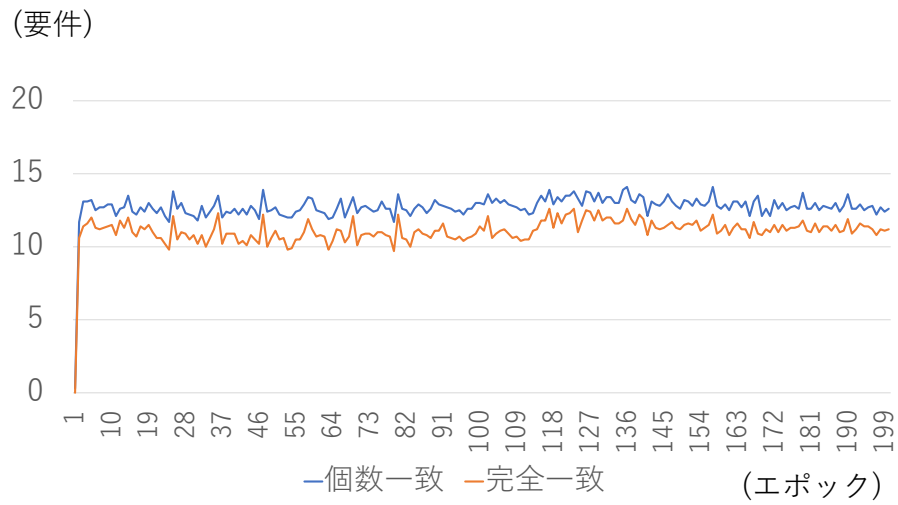


図 21 試験データの要件の個数一致数と完全一致数 (題材 3・SFP 法)

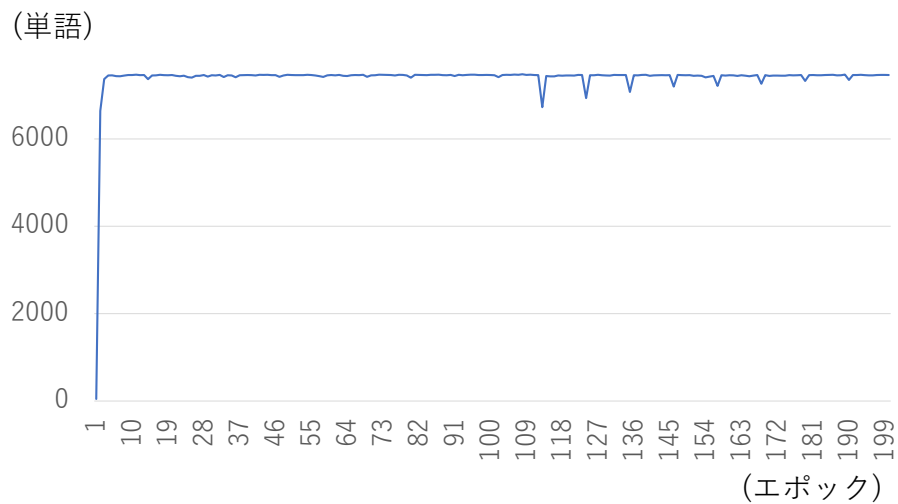


図 22 学習データの単語ごとのラベル一致数 (題材 3・SFP 法)

表 18 学習データと試験データの評価指標の平均 (題材 3・SFP 法)

評価指標	学習データ	試験データ
単語ごとの一致数	7461.2/7552.0	743.9/840.0
要件ごとの完全一致数	184.7/198.0	11.2/22.0
要件ごとの個数一致数	187.3/198.0	12.6/22.0

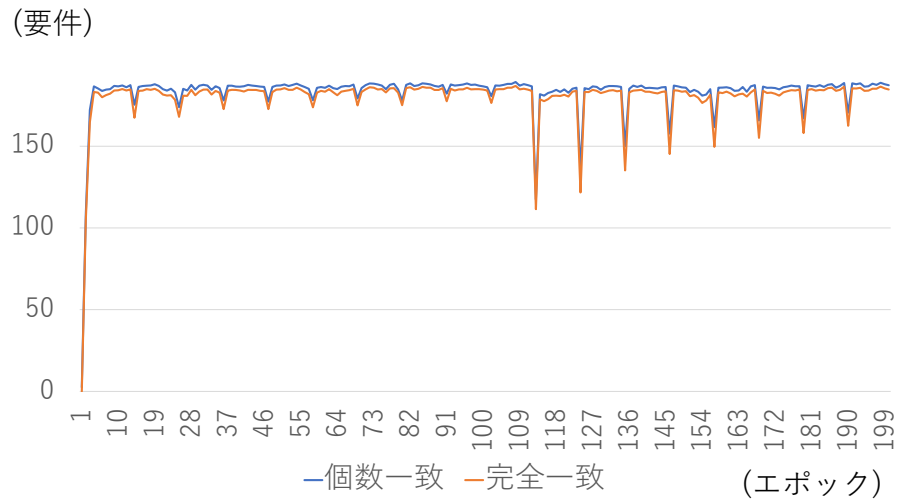


図 23 学習データの要件の個数一致数と完全一致数 (題材 3・SFP 法)

表 19 学習データの統計的指標の平均 (題材 3・SFP 法)

ラベル	F1 値	再現率	適合率
I-item	0.9982	0.9991	0.9974
B-item	0.9992	0.9996	0.9989
I-DF	0.9276	0.9130	0.9480
B-DF	0.9760	0.9699	0.9841
I-TF	0.9026	0.8952	0.9166
B-TF	0.9199	0.9200	0.9271
O	0.9933	0.9935	0.9932

表 20 試験データの統計的指標の平均 (題材 3・SFP 法)

ラベル	F1 値	再現率	適合率
I-item	0.8167	0.8282	0.8260
B-item	0.7675	0.7338	0.8701
I-DF	0.2274	0.2000	0.2750
B-DF	0.3445	0.3283	0.3750
I-TF	0.2094	0.2003	0.2373
B-TF	0.3256	0.3756	0.3752
O	0.9380	0.9439	0.9331

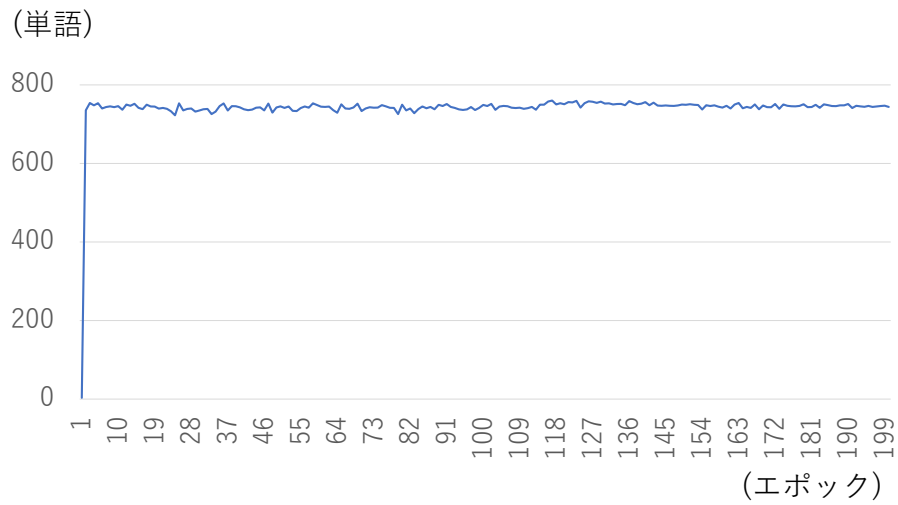


図 24 試験データの単語ごとのラベル一致数 (題材 3・NESMA 法)

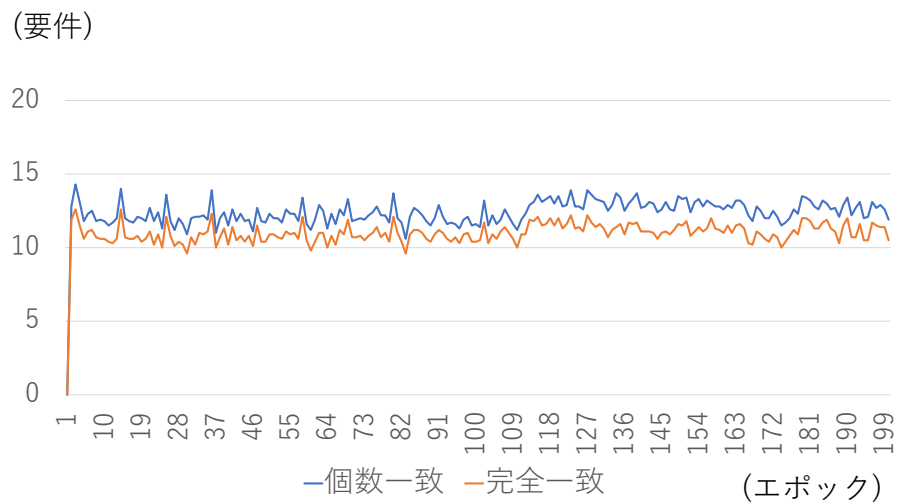


図 25 試験データの要件の個数一致数と完全一致数 (題材 3・NESMA 法)

表 21 学習データと試験データの評価指標の平均 (題材 3・NESMA 法)

評価指標	学習データ	試験データ
単語ごとの一致数	7539.3/7634.8	752.1/849.2
要件ごとの完全一致数	182.7/198.0	10.5/22.0
要件ごとの個数一致数	185.3/198.0	11.9/22.0

表 22 学習データの統計的指標の平均 (題材 3・NESMA 法)

ラベル	F1 値	再現率	適合率
I-item	0.9996	1.0000	0.9991
B-item	0.9998	0.9995	1.0000
I-ILF	0.9227	0.9594	0.8919
B-ILF	0.9587	0.9869	0.9334
I-EIF	0.0000	0.0000	0.0000
B-EIF	0.0000	0.0000	0.0000
I-EI	0.9096	0.8976	0.9240
B-EI	0.9297	0.9327	0.9300
I-EO	1.0000	1.0000	1.0000
B-EO	1.0000	1.0000	1.0000
I-EQ	0.8724	0.8592	0.8949
B-EQ	0.9029	0.9180	0.8955
O	0.9931	0.9924	0.9937

表 23 試験データの統計的指標の平均 (題材 3・NESMA 法)

ラベル	F1 値	再現率	適合率
I-item	0.7555	0.7597	0.7721
B-item	0.6806	0.6961	0.8065
I-ILF	0.1712	0.1933	0.1548
B-ILF	0.2233	0.2381	0.2218
I-EIF	0.0000	0.0000	0.0000
B-EIF	0.0000	0.0000	0.0000
I-EI	0.2821	0.3208	0.3631
B-EI	0.3655	0.4249	0.3734
I-EO	0.0000	0.0000	0.0000
B-EO	0.0000	0.0000	0.0000
I-EQ	0.0167	0.0111	0.0333
B-EQ	0.0471	0.0320	0.0889
O	0.9409	0.9497	0.9325

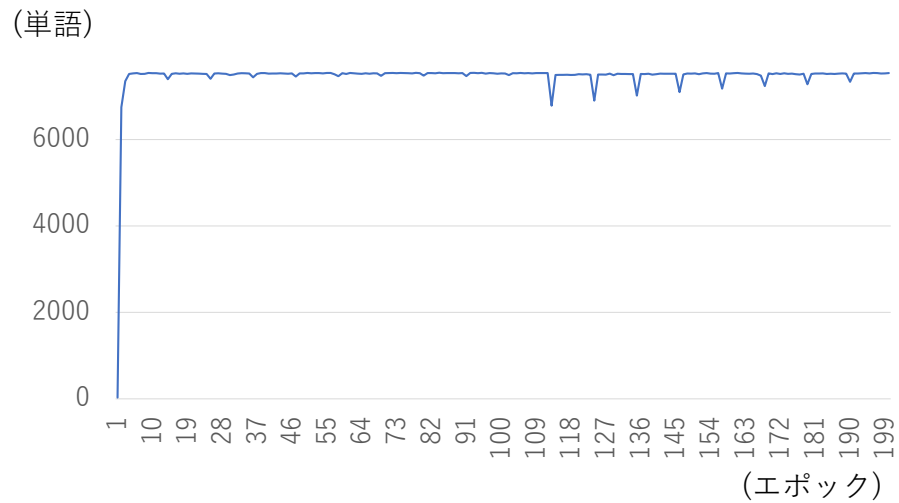


図 26 学習データの単語ごとのラベル一致数 (題材 3・NESMA 法)

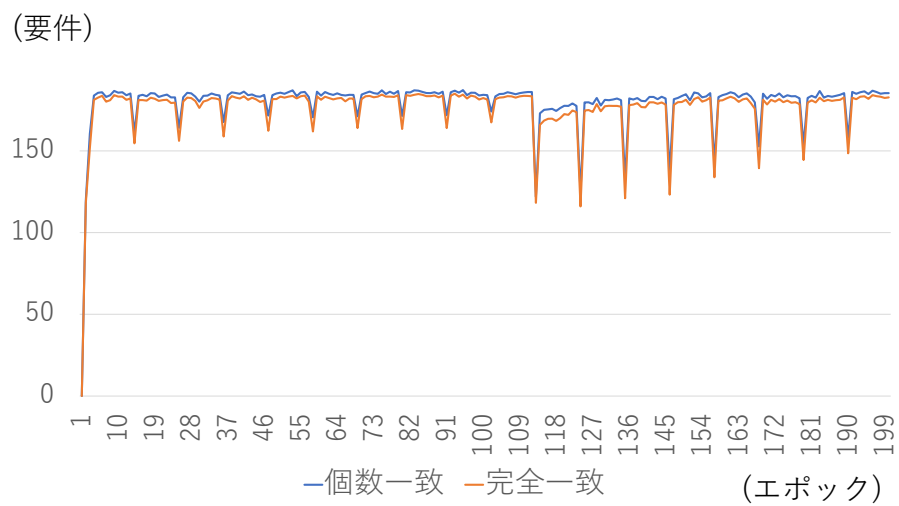


図 27 学習データの要件の個数一致数と完全一致数 (題材 3・NESMA 法)

5 考察

各題材での実験を通して、学習によって学習データに対する基本構成要素の抽出・分類の精度が上がる事が確認できた。試験データに対する基本構成要素の抽出・分類の精度については題材 2 とそれ以外に対する実験で異なる傾向が認められた。この傾向の差は題材から作成したデータによって生まれたものであると考えられる。

題材 2 に用いたデジタル庁の提供するデータ要件・連携要件の標準仕様は記述方式が統一されており、学習に用いた単語数も多い。更に O タグの付与された単語の割合が他の二つの題材より少なく、前後の単語の情報からラベルを予測する本手法での推測が容易であったと考える。

他の題材については、小規模題材は学習データ量が小さく、十分な学習ができなかったと考えられる。

CFPS Fellow によって分類された要件は、文中に表現されない暗黙的な仕様から導かれる要件や、類似する文章であっても要件となりえるものとなりえないものがあるなど、専門家の知見によって抽出・分類される要件が含まれるため、学習結果として反映できなかった要件が多く含まれると考えられる。具体的な例としてログの分析は機能仕様として記述されているものの、外部アプリケーションで分析するため要件の計測を行わなかったり、住民データ出力が機能として明言されている場合、日本人と外国人ではデータ形式が異なるため、暗黙のうちに日本人と外国人で別の要件として扱うなどが挙げられる。また、今回のデータ作成に際しては一つの機能として扱ったが、"管理する"などの一つの単語で複数の基本構成要素を扱う表現もあるため、深層学習のみでは専門家の行う機能の抽出を行う事は難しいと考えられる。

以上より、本手法を用いてより精度の高い抽出・分類を行うためには、記述方式の統一や、仕様の明文化を行う必要があると考える。また、今回の実験では SFP 法を用いて機能仕様書から TF, DF を抽出・分類する場合と、NESMA 法を用いて 5 種類の基本構成要素を抽出・分類する場合での精度は SFP 法を用いた場合が上回るという結果が得られた。しかし、差はわずかであり、異なる題材に適用した場合の精度の違いを確認する必要があると考える。

6 おわりに

本研究では日本語で記述された機能仕様書を対象とした深層学習モデルを用いた FP 計測支援ツールの提案を行った。ツールは既存の機能仕様書から深層学習を行い、自然語で記述されたツールの精度を確認するため、3つの題材を対象に実証実験を行った。一つ目の題材である FP 学習用の教材 2 種を用いて、それぞれ 19 要件と 16 要件から学習することで、それぞれ 4 要件中 1 件、5 要件中 1 件で FP の基本構成要素の分類結果が人間が分類した結果と同様になるという結果を得た。二つ目の題材であるデジタル庁の公開している 33 種の機能仕様書から抽出した要件を用いて 10 分割交叉実験を行い、平均して 215.4 要件中 192.4 要件で FP の基本構成要素の分類結果が人間が分類した結果と同様になるという結果を得た。三つ目の題材である総務省の公開している印鑑登録システムでは、機能仕様書から専門家が抽出した要件を用いて 10 分割交叉実験を行い、平均して 22 要件中 12.6 要件で FP の基本構成要素の分類結果が専門家が分類した結果と同様になるという結果を得た。デジタル庁の公開している 33 プロジェクトの機能仕様書を用いて実証実験を行い、十分な精度を確認した。他方、他の題材では十分な精度とは言えず、十分な精度を出すためにはデータの記述の統一や、要件の明文化が必要であると考ええる。

今後の研究課題として、大規模なデータを用いた際の SFP 法による計測と NESMA 法による計測の精度比較や、異なる題材から学習したモデルの交叉試験を考えている。例として、本研究の題材 2 から学習したモデルを用いて題材 3 から要件を抽出・分類することにより、ユーザーが独自にモデルを作成することの優位性や課題点を明らかにできる可能性がある。加えて、学習の際のエポック数などのパラメーターに関しても最適な値を発見することも今後の課題である。また、FP の算出の為には要件を抽出したのち、同一の機能と見なせる項目に関しては削除を行う必要がある。そのため、FP 算出の完全な自動化の為には他の自然言語処理を用いて、同一の機能を判別する必要がある。

謝辞

本研究を行うにあたり多くの皆様に多大な支援を頂きました。

本研究の実施にあたり、2年にわたりご指導を頂きました楠本真二教授に深く感謝いたします。ファンクションポイントをはじめとするソフトウェア開発の上流行程に関する幅広い知見や、様々な助言を頂きました。研究方針の相談や論文の添削等を根気よく行っていただき、本研究を進めるうえで大きな助けになりました。

本研究への多くの助言、コメントを頂きました、柏本真佑助教に心より感謝申し上げます。中間発表の際に鋭い指摘を頂き、研究をより有意義なものへと昇華することができました。

研究活動を円滑に進められるよう多くのご支援をいただきました、事務補佐員の橋本美砂子氏に心より感謝申し上げます。感染症が猛威を振るう中での研究室生活でしたが、おかげさまで大事なく過ごすことが出来ました。

本研究を行うにあたり、データ作成にご協力いただいた倉重誠氏に深く感謝申し上げます。データのみならず、CFPS Fellow としてファンクションポイントに関わる有意義な議論をできたことに感謝します。

本研究の実施にあたり、助言を頂きました経済調査会の皆様に感謝申し上げます。

研究生活で共に過ごした研究室の皆様には多くの場面で手助けや励ましを頂き、更には多様な物事に関して意見を交わすことで多くの学びを得ることができました。

最後に、本研究の遂行中、叱咤激励して頂いた友人や家族へ、心より感謝の意を示します。

参考文献

- [1] Symons, C.: *Software sizing and estimating: Mk II FPA (function point analysis)*, John Wiley & Sons (1991).
- [2] Reifer, D. J.: Asset-R: A function point sizing tool for scientific and real-time systems, *Journal of Systems and Software*, Vol. 11, No. 3, pp. 159–171 (1990).
- [3] Rask, R., Laamanen, P. and Lyytinen, K.: A COMPARISON OF ALBRECHT’S FUNCTION POINT AND SYMONS’MARK II METRICS (1992).
- [4] Rask, R., Laamanen, P. and Lyytinen, K.: Simulation and comparison of Albrecht’s function point and DeMarco’s function bang metrics in a CASE environment, *IEEE Transactions on Software Engineering*, Vol. 19, No. 7, pp. 661–671 (1993).
- [5] Lamma, E., Mello, P. and Riguzzi, F.: A system for measuring function points from an er-dfd specification, *The Computer Journal*, Vol. 47, No. 3, pp. 358–372 (2004).
- [6] Diab, H., Koukane, F., Frappier, M. and St-Denis, R.: μ ROSE: automated measurement of COSMIC-FFP for Rational Rose RealTime, *Information and Software Technology*, Vol. 47, No. 3, pp. 151–166 (2005).
- [7] Živković, A., Rozman, I. and Heričko, M.: Automated software size estimation based on function points using UML models, *Information and Software Technology*, Vol. 47, No. 13, pp. 881–890 (2005).
- [8] Abrahao, S. and Insfran, E.: A metamodeling approach to estimate software size from requirements specifications, in *2008 34th Euromicro Conference Software Engineering and Advanced Applications*, pp. 465–475IEEE (2008).
- [9] Lind, K. and Heldal, R.: A model-based and automated approach to size estimation of embedded software components, in *Model Driven Engineering Languages and Systems: 14th International Conference, MODELS 2011, Wellington, New Zealand, October 16-21, 2011. Proceedings 14*, pp. 334–348Springer (2011).
- [10] Finnie, G. R., Wittig, G. E. and Desharnais, J.-M.: A comparison of software effort estimation techniques: Using function points with neural networks, case-based reasoning and regression models, *Journal of systems and software*, Vol. 39, No. 3, pp. 281–289 (1997).
- [11] Park, H. and Baek, S.: An empirical validation of a neural network model for software effort estimation, *Expert Systems with Applications*, Vol. 35, No. 3, pp. 929–937 (2008).
- [12] Pai, D. R., McFall, K. S. and Subramanian, G. H.: Software effort estimation using a neural

- network ensemble, *Journal of Computer Information Systems*, Vol. 53, No. 4, pp. 49–58 (2013).
- [13] A. J. Albrecht : *Function point analysis. Encyclopedia of Software Engineering, Vol. 1*, Addison-Wesley Professional, 518-524 pp. (1994).
 - [14] Release 4. 3. International Function Point UsersGroup : IFPUG: Function Point Counting Practices Manual (2010).
 - [15] Standardization (ISO), for I. O.: ISO/IEC 20926:2009 - Software and systems engineering — Software measurement — IFPUG functional size measurement method 2009: <https://www.iso.org/standard/51717.html>.
 - [16] Kusumoto, S., Imagawa, M., Inoue, K., Morimoto, S., Matsusita, K. and Tsuda, M.: Function point measurement from Java programs, in *Proceedings of the 24th International Conference on Software Engineering*, pp. 576–582 (2002).
 - [17] Lavazza, L., Morasca, S. and Robiolo, G.: Towards a simplified definition of Function Points, *Information and Software Technology*, Vol. 55, No. 10, pp. 1796–1809 (2013).
 - [18] Quesada-López, C. and Jenkins, M.: Function point structure and applicability validation using the ISBSG dataset: a replicated study, in *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pp. 1–1 (2014).
 - [19] 山田涼太, 柏本真佑, 肥後芳樹, 楠本真二 : 系統的文献レビューを用いたファンクションポイント研究の調査, 電子情報通信学会論文誌 D, Vol. J103-D, No. 3, pp. 144–158 (2019).
 - [20] IEEE, : IEEE guide to software requirements specifications (1984), <https://chumpolm.files.wordpress.com/2018/09/ieee-std-830-1984.pdf>.
 - [21] 日本情報システム・ユーザー協会 : 要求仕様定義ガイドライン (UVC プロジェクト報告書 2007) (2007), https://juas.or.jp/cms/media/2022/09/uvc_project_report_2007.pdf.
 - [22] NESMA (National Electrical Switchboard Manufacturers Association : オランダソフトウェア計測協会) : Early Function Point Analysis : <https://nesma.org/themes/sizing/function-point-analysis/early-function-point-counting/>.
 - [23] IFPUG (International Function Point USERSGROUP) : Introducing Simple Function Points (SFP) : <https://ifpug.org/2021/10/26/introducing-simple-function-points-sfp>.
 - [24] 柏本隆志, 楠本真二, 井上克郎, 鈴木文音, 湯浦克彦, 津田道夫 : イベントトレース図に基づく要求仕様書からのファンクションポイント計測手法, 情報処理学会論文誌 Vol.41, No. 6, pp. 1895–1904 (2000).
 - [25] Meli, R.: Simple function point: a new functional size measurement method fully compliant with IFPUG 4. x, in *Software Measurement European Forum* (2011).

- [26] 大岩佐和子, 押野智樹, 楠本真二, 松本健一: 開発初期のソフトウェア規模見積りに関する研究—NESMA 概算法の検証—, 経済調査研究レビュー, Vol. 16, pp. 81–88 (2015).
- [27] Grishman, R. and Sundheim, B. M.: Message understanding conference-6: A brief history, in *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics* (1996).
- [28] Khalid, M. A., Jijkoun, V. and De Rijke, M.: The impact of named entity normalization on information retrieval for question answering, in *European Conference on Information Retrieval*, pp. 705–710Springer (2008).
- [29] Liu, T., Zhang, X., Zhou, W. and Jia, W.: Neural relation extraction via inner-sentence noise reduction and transfer learning, *arXiv preprint arXiv:1808.06738* (2018).
- [30] Bowden, K. K., Wu, J., Oraby, S., Misra, A. and Walker, M.: Slugnerds: A named entity recognition tool for open domain dialogue systems, *arXiv preprint arXiv:1805.03784* (2018).
- [31] Huang, Z., Xu, W. and Yu, K.: Bidirectional LSTM-CRF models for sequence tagging, *arXiv preprint arXiv:1508.01991* (2015).
- [32] Wei, H., Gao, M., Zhou, A., Chen, F., Qu, W., Wang, C. and Lu, M.: Named entity recognition from biomedical texts using a fusion attention-based BiLSTM-CRF, *IEEE Access*, Vol. 7, pp. 73627–73636 (2019).
- [33] Zhang, K., Wang, X., Ren, J. and Liu, C.: Efficiency improvement of function point-based software size estimation with deep learning model, *IEEE Access*, Vol. 9, pp. 107124–107136 (2020).
- [34] 山田涼太, 山田悠斗, 楠本真二, 松本真佑, 肥後芳樹: 要件定義書からのファンクションポイント自動計測の試み, ソフトウェア・シンポジウム 2017 (2017).
- [35] JFPUG (Japan Function Point USERSGROUP) : CFPS 概要 : http://www.jfpug.gr.jp/app-def/S-102/wp/?page_id=1374.

付録

7 小規模題材に用いた機能仕様書

7.1 作品管理システム

貸出のため購入した DVD の情報をシステムに登録する

入力する情報は DVD 番号タイトルメーカージャンル枚数である

入力が完了すると DVD 情報と DVD 明細情報が生成される

すべての情報が入力されていない場合はエラーメッセージを出力する

すでにある DVD の枚数を増やす

DVD 番号をキーにして DVD 情報を呼び出す

DVD 番号が誤りならばエラーメッセージを出力する

追加する枚数を入力すると DVD 情報と DVD 明細情報が更新される

DVD 明細情報の枝番は DVD 情報の最大枝番の次の番号から順につけ増枚後の枝番の最大値を DVD 情報の最大枝番に保存する

すでにある DVD の何枚かを処分する

DVD 番号をキーにして DVD 情報を呼び出す

DVD 番号が誤りならばエラーメッセージを出力する

処分する枝番を入力すると該当する DVD 明細情報が削除される

処分対象の枝番が存在しないか貸出中の場合はエラーメッセージを出力する

現在の在庫から条件に合致する DVD の一覧を取り出して表示する

入力できる条件はタイトルメーカージャンルの 3 つ表示される情報は DVD 番号タイトルメーカージャンル蔵枚数貸出可能枚数である

条件に合致する DVD がない場合、その旨を表示する

すべての登録されている DVD の情報を帳票に印刷する

登録されている DVD の情報を抹消する

DVD 番号をキーにして DVD 情報を呼び出す

DVD 番号が誤りあるいはまだ貸出中の枝番がある場合エラーメッセージを出力する

貸出中の枝番があるかどうかは DVD 明細情報に貸出先利用者番号が入っているかどうかで確認する

内容確認後削除を実行すると該当する DVD 情報 DVD 明細情報ともに削除される

7.2 酒屋の在庫管理システム

1. 課題

以下に示す仕様に基づいて、ある酒屋の倉庫における「受付係」の作業を行うプログラムを作成しなさい。

2. 要求仕様

ある酒屋では、すべての在庫（酒）を幾つかのコンテナに載せて、一つの倉庫に保管している。

「受付係」の仕事は、倉庫で在庫の積み下ろしを行う「倉庫係」、及び、客の注文を聞いて客に酒を届ける「配送係」との間で数種類の伝票をやりとりしながら、倉庫内の在庫を管理することにある。

2.1 受付係への入力

(I1) 積荷票

酒を積んだコンテナが倉庫に搬入されると、倉庫係から「積荷票」が送られてくる。

1 台のコンテナに酒を 10 銘柄まで混載できるため、積荷票には、搬入されたコンテナの番号（コンテナに付けられたシリアル番号）、搬入年月、搬入日時と共に、積載銘柄数が記入されている。

さらに、積載されている銘柄と銘柄毎の数量（積載量）が積載銘柄数だけ繰り返し記入されている。

(I2) 出庫依頼票

客から酒の注文があると、配送係から「出庫依頼票」が送られてくる。

出庫依頼票には、出庫すべき酒の銘柄（1 銘柄のみ）、数量、及び、送り先名が記入されている。

2.2 受付係からの出力

(O1) 在庫不足票

出庫依頼票により出庫を依頼されたが、出庫すべき酒が必要量だけ倉庫にないなどの理由により出庫できない場合、配送係に「在庫不足票」が送られる。

在庫不足票には、出庫できなかった酒の銘柄（1 銘柄のみ）、数量、送り先名、及び、対応する出庫依頼票の依頼番号（2.3 の（A1）参照）が記入されている。

(O2) 出庫指示票

出庫すべき酒が必要量以上に倉庫にある場合、倉庫係に「出庫指示票」が送られる。

複数のコンテナから出庫するような指示も許されるので、出庫指示票には、出庫すべき酒の銘柄（1 銘柄のみ）、送り先名、対応する出庫依頼票の依頼番号と共に、出庫の対象となるコンテナの数（最大 10 個）が記入されている。

さらに、コンテナ番号、コンテナ毎の数量（出庫量）、及び、空コンテナマーク（出庫によってコンテナが空になるかどうかを示すマーク）が出庫の対象となるコンテナの数だけ繰り返し記入されている。

2.3 受付係の作業

(A0) 在庫の把握

積荷票や出庫指示票に基づいて、倉庫内のコンテナや酒の在庫状況を常に把握する。

(A1) 出庫依頼票への依頼番号の付加

倉庫係から送られてきた出庫依頼票に、受付順に 1 からの通し番号を依頼番号として付加する。

依頼番号の付加は、出庫依頼票が送られてきた時点で行う。

(A2) 在庫不足票の作成

出庫依頼票に依頼番号を付加した時点で、出庫依頼票で出庫するよう指定された酒が必要量だけ倉庫にない場合、在庫不足票を作成し、配送係に送ると共に、その控え（「在庫不足票控」）を作成し、保管する。

ただし、指定された酒と同一銘柄の酒に対する在庫不足票控が存在する場合には、在庫の有無に関わらず在庫不足票を作成し、配送係に送ると共に、その控えを作成し、保管する。

なお、在庫不足票控は在庫不足票と同一形式で、その記入内容是对應する在庫不足票と同じとする。

(A3) 出庫指示票の作成

出庫依頼票に依頼番号を付加した時点で、出庫依頼票で出庫するよう指定された酒と同一銘柄の酒に対する在庫不足票控が存在せず、かつ、指定された酒が必要量以上に倉庫にある場合、出庫指示票を作成し、倉庫係に送る。

また、在庫不足票控が存在する場合には、新たに在庫が搬入された（積荷票が倉庫係から送られてきた）時点で在庫を確認し、在庫不足票控で指定された酒が必要量以上に倉庫にある場合、出庫指示票を作成し、倉庫係に送る。

なお、出庫指示票の作成においては、次のような制約条件がある。

- 出庫の対象となるコンテナが複数の場合には、コンテナ番号の小さいものから順に記入されているものとする。
- 出庫の対象となるコンテナ数が 10 を越える場合には、コンテナ番号の小さいものから順にコンテナを 10 個ずつのグループに分け、グループ毎に出庫指示票を作成するものとする。
- 出庫すべき酒が必要量よりも多く倉庫にある場合には、コンテナの搬入年月、及び、搬入日時に基づき、古いものから出庫されるように出庫指示票を作成するものとする。
- 在庫不足票控が複数存在し、かつ、新たな在庫搬入によって同時に複数の出庫指示票が作成される場合には、対応する在庫不足票控の依頼番号の小さいものから順に出庫指示票を作成するものとする。
- 在庫不足票控で指定された酒が必要量以上に倉庫にある場合であっても、その依頼番号よりも小さな依頼番号を持つ同一銘柄に対する在庫不足票控が存在する場合には、出庫指示票を作成することはできないものとする。

(A4) 在庫不足票控の破棄

在庫不足票控に基づいて出庫依頼票を作成し、倉庫係に送った場合には、対応する在庫不足票控をその時点で破棄する。

3. 外部仕様

プログラムへの入力「積荷票」と「出庫依頼票」の時系列であり、出力はそれに基づいて作成された「在庫不足票」と「出庫指示票」の時系列である。

プログラムの実行ファイル名は sakaya で、実行方法は次の通り。

なお、実行開始時点では、倉庫内に在庫はないものとする。

sakaya への入力は UNIX の標準入力から与えられる。

入力には構文的にも意味的にも誤りは含まれないものとする。

なお、入力をファイルから与える場合には、そのファイル名の拡張子は test とする。

sakaya の出力は UNIX の標準出力に対して行う。

4. sakaya の入出力データフォーマット

[入力データフォーマット]

積荷票… 票識別子、コンテナ番号、搬入年月、搬入日時、銘柄数からなる 1 つのレコードと、銘柄、数量からなる 1 つ以上のレコードで構成される。

「搬入年月」は、上位 2 桁が西暦年の下位 2 桁を、下位 2 桁が月（1～12）を、それぞれ表す。

「搬入日時」は、上位 2 桁が日（1～31）を、下位 2 桁が時（0～23）を、それぞれ表す。

出庫依頼票… 票識別子、送り先名、銘柄、数量からなるレコードである。

[出力データフォーマット]

在庫不足票… 出力メッセージからなる 2 種類のレコードと、依頼番号、送り先名、銘柄、数量からなる 1 つのレコードで構成される。

注) 各データ項目は 1 つ以上の空白で、各レコードは 1 つの改行記号で、各伝票は 1 つの改行記号で、それぞれ区切られるものとする。