

共起関係を用いて変更要求からプログラムへの影響分析を行う手法の提案と評価

川畑徹平^{†1} 岩崎陽也^{†1} 中島毅^{†1} 徳本修一^{†2} 塚本良太^{†2} 高橋加寿子^{†2}
芝浦工業大学^{†1} 三菱電機株式会社情報技術総合研究所^{†2}

1. はじめに

ソフトウェア影響分析は、変更要求を実現する際、その影響が及ぶ成果物範囲を決定する作業である[1][2]。影響範囲を見逃すと、変更要求の不完全な実装や既存機能への悪影響のテスト漏れに繋がる。特に、大規模な流用元プログラムに対して小規模な流用開発を行うプロジェクトにおいて、影響分析の精度と効率は開発の生産性と品質を決定づける主要因となる[3]。

変更要求から流用元プログラムの修正対象を特定する影響分析を行う場合、修正候補の絞り込みと修正対象の決定の2つのタスクが必要になる。後者のタスクは開発者自身が修正候補をレビューすることにより行うしか方法がないため、前者のタスクをいかに抜けなくかつ無駄なく実施できるかが影響分析の精度と効率を決定づける。

従来、修正候補の絞り込みをするためには、要求トレーサビリティを用いることが一般的である。要求トレーサビリティは、要求と関連する要求、実装、および検証との間の識別可能な成果物間の関連性である[4]。しかし、トレーサビリティの確立と維持を高い精度で継続し続けることは一般に困難である。また、そもそも変更要求から既存のトレーサビリティを利用するためには、まず影響を受ける要求を正しく識別する必要がある。

我々は、漏れなく無駄の少ない修正候補の絞り込みを達成するために、変更要求に対する変更履歴を機械学習することにより、変更要求文から直接流用元プログラムのコンポーネントリストを修正候補として推定する手法を提案し実装を行った[5][6]。この手法は、変更要求文をベクトル化し、そのベクトルを入力、コンポーネント毎の修正の有無を出力として機械学習および推定する。

論文[5][6]では、提案手法の機械学習部分の実装として、畳み込みニューラルネットワーク(CNN)を採用している。この実装を実データにより評価することにより、大規模な流用元プログラムに対して小規模な流用開発を行う多数プロジェクトがある状況下において効果的であることを示した。

しかしながら、修正候補を sigmoid 値の高い順に並べるランキング方式で上位に修正対象が密度高く現れることは確認できたが、実際にレビュー作業を行う対象プログラムを絞り込むためには、修正候補の範囲を決定する必要がある。

た。そのため実データから閾値を変化させ対象範囲を30%に絞った場合の閾値を求め修正候補の絞り込みを行ったところ修正候補から外れる修正対象の比率(ハズレ率)は23%となり、このままでは修正対象の見逃しにつながるため改善が求められる結果となった。

我々は、提案手法の実装方法のハズレ率を改善するため、流用元プログラムのコンポーネント間の依存関係に注目し、依存関係の存在するコンポーネントは同時に修正されることが多いという仮説を立てた。この仮説が正しければ、同一変更要求に対して依存関係のある修正コンポーネントは共起しやすいことになる。

本論文は、我々は、提案手法の機械学習(多ラベル分類)に共起関係を考慮したアルゴリズムを採用しその効果を調べることでこの仮説の検証を行う。この目的のため、アルゴリズムとして CNN 法、BR (Binary Relevance) 法、LP (Label Powerset) 法、RAkEL (Random k-Labelsets) 法を適用し、ハズレ率と正答率の比較評価を行った。その結果 CNN による実装よりも RAkEL 法の推定精度が 7.4 ポイント向上した。本論文は、2章で論文[5][6]で示した提案手法とその実装方法について示し、3章で多ラベル分類およびその共起関係を考慮するアルゴリズムを紹介し、4章で各実装の比較を行い結果の分析結果を示す。

2. 提案手法とその CNN による実装

2.1 対象プロジェクトとその成果物の特徴

多品種かつ継続的に進化する製品のソフトウェア開発(多品種少変更開発)[7]では、比較的大規模なソースコード基盤に定期的に新機能を追加し、異なるハードウェアセットや様々な出荷先に合わせてカスタマイズするために、多数の変更が行われる。このような開発では、複数の変更要求を持つ小規模なプロジェクトが多数並行し、影響度分析を正確かつ効率的に行えるだけのソースコード基盤の知識を持つ人材が不足することが多い。

対象プロジェクト群は、XDDP[8]と呼ばれる派生開発手法を採用しており、変更要求ごとに1つの変更設計書が作成される。この文書には以下の項目が記述されている。

- 変更要求 ID
- 要求事項（自然言語）
- 実装
 - ソフトウェア方式設計仕様の変更内容
 - ソフトウェア詳細設計仕様のモジュールに関する変更内容
 - ソースコードの変更内容（変更したコンポーネントやモジュールの名称を含む）

毎年約 30 件のプロジェクトが発生し、1 件あたり平均 10 件程度の変更要求が実施されており、提案手法適用時までには、1000 件近くの変更設計書が作成されていた。

学習の入力となる要求事項は日本語テキストで 20 から 400 文字、出力となるソースコードベースのコンポーネント数は 32 個である。

2.2 提案手法

提案手法は、大量の変更設計書を学習し、新しい変更要求から直接から影響範囲となる修正候補のコンポーネントを推定する [5][6]。

図 1 に示すように、学習時には 1 つの変更設計書に対して、まず変更要求文を抽出し文書ベクトル化技術を用いて変更要求文のベクトルを作成する。次に、その変更要求文に対応するソースコードの修正箇所の情報からコンポーネントベクトルを作成する。このベクトルはコンポーネント番号(1-32)のインデックスの値に 0 か 1（1 が修正あり、0 が修正なし）を入れたものになっている。

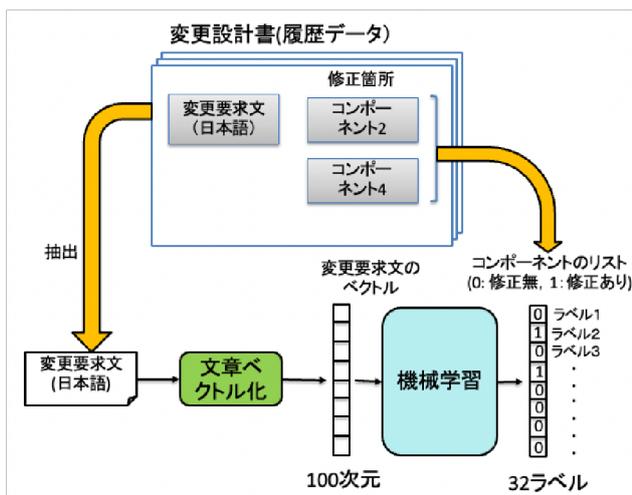


図 1 提案手法の機械学習の構成法

図 2 に示すように、提案手法は、推定時において、新たな変更要求文が入力された時に作成した学習モデルを用いて修

正候補のコンポーネントのリストを修正される可能性の高い順に出力する。

提案手法は、トレーサビリティを用いる影響分析に比べて、開発活動に負担をかけることなく、新しい変更要求文から直接修正候補を絞り込むことができる利点を持っている。



図 2 提案手法による修正候補推定法

2.3 先行研究のアルゴリズム構成

(1) 文章ベクトル化

先行研究では、文章をベクトル化する際、まず文章を形態素解析器 mecab にかけて語に分解する。分解された語に対して、その中から利用する語の抽出（語抽出）、語のベクトル化、ベクトル統合の 3 段階の処理を行う。結果としての文章ベクトルは 100 次元とした。

先行研究では、この各段階の実装候補を組み合わせた 3 つの実装方法（表 1）を適用し推定精度を出したところ名詞選択 + doc2vec(実装 3) が最もよい結果を出すことが確認できた。

表 1 先行研究における文章ベクトル化の実装方法

実装 No	語抽出	語のベクトル化	ベクトル統合
実装 1	名詞抽出	word2vec (skip-gram)	単純平均
実装 2	全選択		doc2vec
実装 3	名詞抽出		doc2vec

(2) 機械学習

変更要求文のベクトルに対して、32 個のベクトルの値の 0 か 1 かを決定するため、機械学習の多ラベル分類問題と見なすことができる。

先行研究では、この多ラベル分類問題の機械学習法として畳み込みニューラルネット (CNN) を用いて実装している。

図 3 に実装された CNN の構成を示す。入力の変更要求文をベクトル化した 100 次元のベクトル、出力はコンポーネントリストである 32 次元のベクトルである。

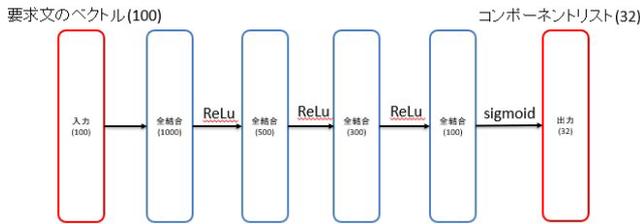


図 3 先行研究の CNN の構造

CNN のパラメータは以下の通りとなっている。

- ・ 中間層: 4 (1000, 500, 300, 100)
- ・ エポック数: 50
- ・ バッチサイズ: 50
- ・ 学習率: 0.1
- ・ 出力時関数: sigmoid

2.4 先行研究の再評価と課題点

先行研究では、変更要求文に対して変更される可能性の高い順にコンポーネントのリストを並べ、影響分析結果として出力する。出力のリストには正解のコンポーネントと不正解のコンポーネントが混在している。正解のコンポーネントとは変更要求によって修正が必要なコンポーネントを指し、修正が必要ないコンポーネントを不正解としている。

先行研究の評価方法は最後に正解のコンポーネントが現れる位置までを網羅範囲比率、その位置までの正答率を網羅範囲正答率の 2 つを評価値として用いており、ランキングの上位 20% の範囲を探せば全ての修正対象を含み、候補の 60% 以上が正答になるというよい結果を得ている。

しかし、先行研究では、実際にレビューしなければならないコンポーネント範囲を決定する方法を提供していない。そこで、本論文では、sigmoid 値に閾値を設けてその閾値以上のコンポーネントの修正候補を決定する方法を実装した。さらに、影響分析に対する効果を適切に評価するために、図 4 に示す 3 つの測定量を定義し測定した。

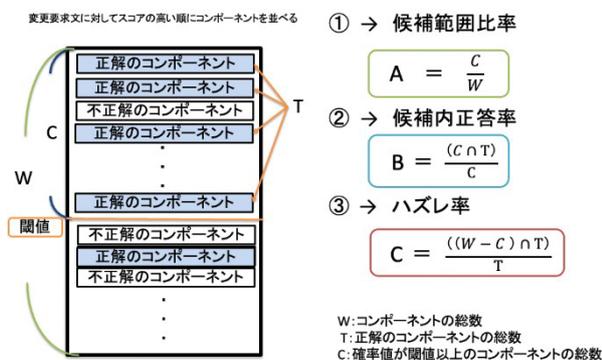


図 4 影響分析への効果を評価するための測定量

- ① 候補範囲比率：出力される sigmoid 値から閾値を設定して閾値以上のコンポーネントが占める比率
- ② 候補内正答率：閾値以上のコンポーネントの内、正解のコンポーネントが占める割合。
- ③ ハズレ率：全ての正解のコンポーネントの内、閾値以下のコンポーネント内に存在する正解のコンポーネントの割合。

閾値を変化させて候補範囲比率が 30% になる時点での、候補内正答率とハズレ率を、先行システムに対して測定した結果を表 2 に示す。

表 2 先行研究の精度

閾値	候補範囲比率	候補内正答率	ハズレ率
0.06	30.0%	35.0%	23.0%

先行システムの機械学習法では候補範囲比率 30% 付近で、候補内正答率が 35%、ハズレ率が 23% という結果となった。問題はハズレ率の高さである。ハズレ率の高さはレビュー漏れの発生のしやすさとなるため、直接的にバグやエラーにつながる。そのためハズレ率の低減が実用化には必要不可欠であり、改善方法を考えることが必要である。

2.5 改善目標

本研究では、候補範囲比率を 30% 以下かつハズレ率を 5% 以下に下げることが目標とした。

レビューの観点で修正候補範囲を絞り込んで効果のあるのが 30% 以下であること、ハズレ率 5% 以下としたのは品質保証の観点から見逃しを 2σ 区間に抑え込みたいためである。

3. 多ラベル分類および共起関係を考慮するアルゴリズム

3.1 共起関係とその着目理由

本研究では、提案手法の実装方法のハズレ率を改善するため、流用元プログラムのコンポーネント間の依存関係に注目し、依存関係の存在するコンポーネントは同時に修正されることが多いという仮説を立てた。この仮説をもとに機械学習アルゴリズム改善に取り組む。この仮説の根拠として、アーキテクチャ的な依存関係

- ・ 関数の呼び出し関係
- ・ リソースの共有関係（通信、メモリ、I/O）
- ・ ファイルの読み書き関係

- 外部関数の読み書き関係
- ファイルのインクルード関係

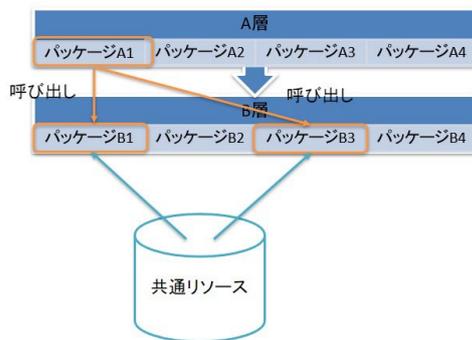


図 5 アーキテクチャから生じる依存関係特に、対象プロジェクトのコンポーネントは、図 5 に示すような階層構造を持っており、A 層のコンポーネントと B 層のコンポーネントには決まった呼び出し関係があること、特定の層(図 5 では B 層)は共通のリソースを扱っており間接的な依存関係が存在することなどがある。

3.2 多ラベル分類を扱うアルゴリズム

先に述べたように、本研究における機械学習は、文章ベクトルの入力集合を、修正を要する複数のモジュールに対応付ける多ラベル分類の問題に帰着する。多ラベル分類とは、一つの対象に対して複数のラベルが割り当てられる分類のことである[9]。多ラベル分類がシングルラベル分類と大きく異なる点は、予測の際にラベル間の共起関係を推定に生かすことで精度改善を見込める点である。そのため本研究では、出力間の共起関係を組み込むために、多ラベル分類を扱うためのアルゴリズムと分類器には Support Vector Machine(SVM)を組み合わせるアプローチを採った。

SVMとは、自然言語処理や音声認識など、分類と回帰の問題に使用することのできる教師あり学習アルゴリズム[10]のことである。今回、分類器に SVMを利用した理由としては SVMが各種の問題に適用されて優秀な成績をおさめており、精度の高い学習アルゴリズムであるため本研究の精度向上につなげることができると考えたからである。

多ラベル分類におけるラベルの予測手法として多ラベル分類をシングルラベル分類に変形させてシングルラベル分類器を用いて予測する方法がある。多ラベル分類の代表的な手法の1つに Binary Relevance法(以下 BR法)がある。

BR法は、個々のラベルについて2値分類器を作成し、分類器の結果の和集合を出力する手法である[11]。BR法は計算量が小さいという利点がある一方で、各ラベルの推定

を個々に行うためラベル間の共起関係を考慮できないという欠点を持つ。本研究では BR法の学習アルゴリズムの構成として、入力文をベクトル化した特徴量に対してコンポーネントを 1:1 で学習させた(図 6)。

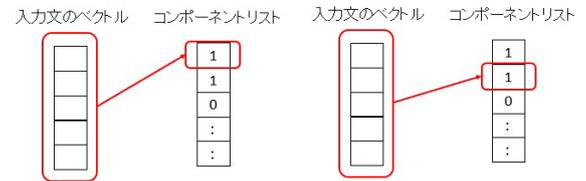


図 6 BR法の学習アルゴリズム構成

3.3 共起関係をモデル化するアルゴリズム

(1) Label Powerset法(LP法)

LP法はラベルの冪集合をクラスとして扱い、マルチクラス分類問題と変換する。冪集合とは考えられる全ての組み合わせのことであり、例えば 1, 2, 3 のラベルの冪集合は、1, 2, 3, 1&2, 1&3, 2&3, 1&2&3 という組み合わせをそれぞれラベルとして扱うことを指す。つまり、マルチラベル分類からマルチクラス分類への変換とは、データ内で見つかった一意のラベルの組み合わせをクラスとして分類することで、1つの入力に対して1つのクラスが定まるマルチクラス分類へと変換できることを指し、

LP法はそのクラスごとに学習を行う(図 7)。

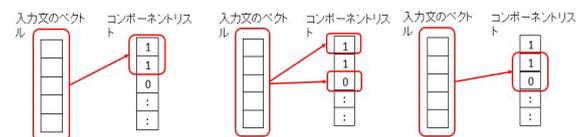


図 7 LP法の学習アルゴリズム構成

その後、その後データに対する各クラスの出現確率を計算し、出現確率をラベルごとに重みづけを行った後、総和を集計する手法(図 8)のことである。

モデル	クラスの出現確率	ラベル別推定結果			
		ラベル1	ラベル2	...	ラベル31
クラス1分類器	0.1	1	0	...	0
クラス2分類器	0.2	0	0	...	1
⋮	⋮	⋮	⋮	⋮	⋮
クラスN分類器	0.3	1	0	...	-
result	result	0.1*1+0.3*1	0	...	0.2*1

図 8 各ラベルの推定結果

LP 法はラベル間の共起関係を利用した予測ができるという利点がある一方で、計算量がラベル数に対して指数関数的に増大してしまうという欠点をもつ。以上から、BR 法と LP 法の間には、“計算量”と“ラベル間の共起関係を利用することによる精度向上”の二つの間のトレードオフ関係があることがわかる。

(2) Random k Labelsets 法(RAKEL 法)

LP 法は、先ほど述べた計算量がラベル数に対して指数関数的に増大してしまうという欠点と別に学習時で観測されたラベル集合しか予測できないため、予測時の新しいラベル集合は予測できず精度が上がらざらという欠点も存在する。前述の LP 法の課題点を解決するために RAKEL 法 [12]が提案された。

図 9 に示すように RAKEL 法は、まず入力時のラベル集合に対してランダムにいくつかの小さなサイズ k のラベル部分集合を作り、その部分集合ごとに LP 法を適用する。

その後、図 10 に示すように、部分集合毎の分類結果を統合してラベルを予測する。

RAKEL 法は、部分集合の LP 計算が全集合に対するそれ (LP 法) と比較して大幅に計算量が削減でき、また各クラス値の分布の偏りも少なくすることができるため LP 法と比較して精度向上の可能性が高いとされているアルゴリズムである。

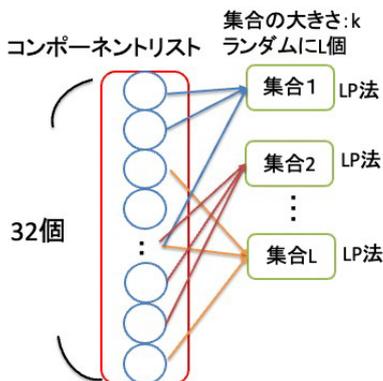


図 9 ラベル集合から部分集合を作る

モデル	ラベル別推定結果				
	ラベル1	ラベル2	...	ラベル31	ラベル32
集合1分類器	1	-	...	0	-
集合2分類器	-	0	...	1	-
:	:	:	:	:	:
集合L分類器	0	0	...	-	1
result	T_1/M_1	T_2/M_2	...	T_{31}/M_{31}	T_{32}/M_{32}

T1: 推定結果が1であるセル数, Mi: 推定結果を持つセル数

図 10 ラベル別推定結果

問題領域の性質や利用可能な学習データ数により、これらのアルゴリズムの適用可能性や精度に違いが出るものとして考える。本研究では、提案手法の機械学習部分に、BR 法、LP 法、RAKEL 法を適用し、同じプロジェクトデータを用いて精度比較を行う。

4. 評価実験

4.1 実験目的

本実験は、共起関係を考慮することで精度が向上するかどうかを検証することを目的とする。

ここで、BR 法、LP 法、および RAKEL 法のアルゴリズムのベースとなるシングルラベルの分類器として、少ないデータ数でも比較的高い精度の出すことができる Support

Vector Machine(SVM)を利用することとした。

まず比較対象として、共起関係を考慮していない BR 法と SVM を組み合わせた手法の精度を測定する。次に、共起関係を考慮した LP 法および RAKEL 法と SVM の組み合わせた手法の精度を測定する。これにより、共起関係による精度が向上するかどうかを評価する。さらに CNN を用いた分類器の測定結果を示すことで、従来研究からの精度改善が

目標レベルにあるかどうかを評価する。

4.2 実験方法

実験は以下の手順で行った。

1. 変更要求文を名詞抽出+doc2vec によりベクトル化し特徴量を抽出。
2. 機械学習コンポーネントを、Scikit-learn を用いて構成。多ラベル分類器として BR 法/LP 法/RAKEL 法 (k=3) +SVM の 3 つを実装。

3. 4.2に示す実験データ 5セットに対して、以下を実施し
平均をとって精度を算出
(ア) 学習データから学習モデルを作成.
(イ) 残りのテストデータでラベルの推定を実施.

4.3 実験データ

本実験では共同研究先である三菱電機株式会社情報技術総合研究所から提供を受けた 405 冊の変更設計書のデータを用いた. 変更設計書を学習用とテスト用に 4:1 の比率で分け, 324 冊を学習用の訓練データに, 81 冊をテストデータとして実験を行う(図 11). また,この一連の流れを他の組み合わせで 5 回行い結果はその平均を実験結果として算出した.

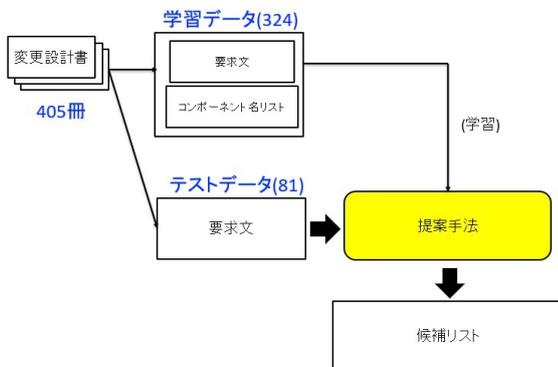


図 11 評価実験のデータ

4.4 実験結果の評価方法

評価方法については, レビュー範囲を示す候補範囲比率, レビュー範囲に存在する修正が必要なコンポーネントの割合を示す候補内正答率, 全ての修正が必要なコンポーネントの内, レビュー範囲外に存在する修正が必要なコンポーネントを示すハズレ率の 3つの測定量を求めた. 0.01 刻みで SIGMOID 値の閾値を動かし, 候補範囲比率が 30%に一番近くなる時を選び各測定量の比較を行った.

4.5 実験結果

本実験では先行研究に合わせて SIGMOID 値を 0.04 から 0.1 の範囲で計測を行なった. また, 候補範囲比率が 30%に一番近い時の評価値を赤字で示している.

(1) BR 法+SVM

表 3 BR 法+SVM の評価値

閾値	候補範囲比率	ハズレ率	候補内正答率
0.04	38.10%	12.30%	14.70%
0.05	33.50%	15.20%	17.10%
0.06	29.40%	17.10%	19.10%
0.07	25.70%	20.50%	21.00%
0.08	23.10%	23.70%	22.40%
0.09	21.10%	28.10%	23.10%
0.1	19.40%	29.90%	24.50%

(2) LP 法+SVM

表 4 LP 法+SVM の評価値

閾値	候補範囲比率	ハズレ率	候補内正答率
0.04	46.00%	10.00%	16.80%
0.05	39.90%	13.40%	18.70%
0.06	35.80%	16.70%	19.90%
0.07	32.80%	19.70%	21.00%
0.08	29.70%	23.00%	22.20%
0.09	26.60%	26.40%	23.70%
0.1	24.10%	29.90%	25.00%

(3) RAKEL 法+SVM

表 5 RAKEL 法+SVM の評価値

閾値	候補範囲比率	ハズレ率	候補内正答率
0.04	41.40%	9.60%	18.70%
0.05	36.50%	11.30%	20.80%
0.06	32.60%	13.50%	22.70%
0.07	29.50%	15.60%	24.50%
0.08	27.20%	16.70%	26.20%
0.09	25.20%	18.80%	27.60%
0.1	23.10%	20.90%	29.30%

(4) 全手法の比較結果

表 6 は候補範囲比率 30%付近で各手法の精度比較を行なった表である. 表 6 候補範囲比率 30%付近の各評価値

手法	閾値	候補範囲比率	ハズレ率	候補内正答率
CNN(従来研究)	0.06	30.0%	23.0%	18.0%
BR+SVM	0.06	29.4%	17.1%	19.1%
LP+SVM	0.08	29.7%	23.0%	22.2%
RAKEL+SVM	0.07	29.5%	15.6%	24.5%

以下の分析結果を得た.

- 従来研究(CNN)と BR 法+SVM を比較した場合ハズレ率の精度が 5.9%向上している. この結果からこの分類問題に対して, SVMは CNN よりも精度が高いことがわかった.
- BR 法と LP 法を比較した場合, LP 法は, BR 法と比較して精度が下がっており, 共起関係を考慮した手法

に優位性が現れるという予想された結果は得られなかった。この原因として、出力のラベル数が 32 個と多いためその組み合わせ数も膨大になってしまったことによる過学習[13]の可能性が高い。

- RAKEL 法+SVM は上記の手法の中でも最も精度が高く BR 法と比較してハズレ率の精度が 1.5%、候補内正答率が 5.4%向上した。過学習が生じなかったものとする。この結果より、共起関係を考慮することにより精度向上に一定の改善効果があるということは確認することができた。
- 候補範囲比率 30%付近のハズレ値は、最も精度の高い RAKEL 法で 15.6%であり、目標精度である 5.0% 以下は達成できなかった。

5. まとめと今後の課題

本論文、では影響分析の精度向上のために共起関係を考慮するアルゴリズムの実装と従来研究との比較を行った。実際に提案手法の分類器として適用した結果、従来の CNN に比べて、共起関係を考慮した RAKEL 法は 7.4 ポイントの精度向上することが分かった。

今後の課題としては、RAKEL 法の改善アルゴリズム (overlapping) の適用による精度改善と、別のデータセットへの適用による提案手法のさらなる有効性の確認と目標精度を達成するための履歴データの必要規模の模索が挙げられる。

参考文献

- [1] Sunil Sikka and Ankit Dhamija: Software Change Impact Analysis, BookRix (2020)
- [2] Bohner: Impact analysis in the software change process: a year 2000 perspective, 1996 Proceedings of International Conference on Software Maintenance, 1996, pp. 42-51, DOI: 10.1109/ICSM.1996.564987.
- [3] H. O. Ali, M. Z. A. Rozan and A. M. Sharif: Identifying challenges of change impact analysis for software projects, 2012 International Conference on Innovation Management and Technology Research, 2012, pp. 407-411, DOI: 10.1109/ICIMTR.2012.6236428.
- [4] 宇多川佳久,他.情報システム標準におけるトレーサビリティの事例と今後.情報処理学会,第 51 巻第 2 号:pp150-158(2010)(参照日 2022-01-01)
- [5] 岩崎 陽也, 中島毅.過去の変更仕様書を学習することにより要求文からプログラム変更部分の候補を推薦するシステムの提案. 情報処理学会 研究報告ソフトウェア工学 (SE), 2021.16, pp. 1-6 (2021).
- [6] Haruya Iwasaki, et al.: A Software Impact Analysis Tool based on Change History Learning and its Evaluation, ICSE-SEIP '22, May 21–29, 2022, Pittsburgh, PA, USA.
- [7] Nagamine, Motoi, Tsuyoshi Nakajima, and Noriyoshi Kuno: A case study of applying software product line engineering to the air conditioner domain. Proceedings of the 20th International Systems and Software Product Line Conference. 2016, pp. 220-226, DOI: 10.1145/2934466.2934489.
- [8] 清水 吉男,「派生開発」を成功させるプロセス改善の技術と極意. 技術評論社 (2007)
- [9] 吉村卓亮,馬場雪乃, and 鹿島久嗣. "相互依存モデルによるマルチラベル分類." 人工知能学会全国大会論文集 第 31 回全国大会, 人工知能学会全国大会論文集:pp. 2K21in1-2K21in1 (2017).
- [10] 山下浩, and 田中茂. "サポートベクターマシンとその応用", (株)数理システム, (2001).
- [11] Tsoumakas, G. and Katakis, I.: Multilabel classification: An overview, Int J Data Warehousing and Mining, Vol. 2007, pp.1-13 (2007)
- [12] Read, J., Puurula, A., & Bifet, A. Multi-label classification with metalabels. In 2014 IEEE international conference on data mining (pp. 941-946). IEEE. (2014, December).
- [13] 鈴木大慈. "過学習と正則化." 応用数理学会第 28 巻,第 2 号: pp28-33. (2018)