

平成 30 年 6 月 26 日現在

機関番号：34412

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K01100

研究課題名(和文)プログラムの説明概念整理に基づくプログラム相互理解向上に向けた教材開発

研究課題名(英文) Developing a programming concept dictionary and designing teaching material based on the concept to enhance mutual understanding of program

研究代表者

竹内 和広 (Kazuhiro, Takeuchi)

大阪電気通信大学・情報通信工学部・教授

研究者番号：20440951

交付決定額(研究期間全体)：(直接経費) 1,900,000円

研究成果の概要(和文)：本研究では、プログラム教育を視野に、プログラムの相互理解や説明の上で必要となる概念を言語表現の観点から整理するプログラム説明概念辞書構築の研究を行った。具体的には、辞書の知識表現に計算機による二次利用が可能な表現を採用し、プログラム説明やプログラム事例を分析するマイニング技術を開発し、知識整理を有効に支援する形で辞書構築を行った。この概念辞書構築により、プログラム教員において、プログラム理解の説明をより客観的に分析することが可能となった。また、プログラム学習者の習熟度合いに応じた段階的な教材の導入・構成法とその指導に資する知見を得ることができた。

研究成果の概要(英文)：In this study, we developed a concept dictionary to organize the concepts necessary for realizing mutual understanding of computer programs through natural language. We developed mining techniques to analyze explanatory examples of programs and program repositories. In the dictionary, we adopted machine-readable expression of words for explanatory concepts. By constructing a dictionary, programming teachers can more objectively analyze learners' understanding through the explanatory descriptions of the programs. In addition, we confirmed that the results of this study can contribute to the development of teaching techniques and methods for designing teaching material.

研究分野：情報工学

キーワード：教育工学 情報教育 プログラム教育 ソフトウェア工学

1. 研究開始当初の背景

大学生の国語力の低下が指摘される中、通常の文章の読み書きだけでなく、抽象的な対象を表現するプログラムを言葉から理解した上で想像・作成し、また、プログラムの理解を言葉で説明する力の低下が顕在化していることが研究背景となった。

例えば、高度なソフトウェア技術者には、ソフトウェア構成の基礎となる実装パターンに関する理解と説明力は前提であり、近年のコミュニケーションを基盤とするアジャイル開発でも、ますます重要性を増す基盤的能力であり、この力の欠如は、高度化する情報産業の中でのプログラム開発モデルにおいて、工程間のコミュニケーションの障害となり、生産性と製品品質を低下させる危険性をもつ課題であった。また、オフショア開発における日本人と海外技術者とのやり取りや、国際的なオープンソースプロジェクトなどでは、この力の欠如が障害となる可能性も高く、日本人の高度なソフトウェア構築への参画には、この力の錬成・向上が必須となるため、教育的に重要な課題でもあった。

研究代表者は、このような課題の要因として、プログラムの理解伝達の際に、プログラム作成者の暗黙的なプログラム理解と、そのプログラムを読んだ第三者のプログラム理解との食い違いに着目した。そして、その原因を解消するためには、プログラム作成者やプログラムの読む手がプログラムの理解を自然言語で的確に伝えることが必須であり、その言語表現による説明を客観的に評価することが重要と考えた。さらに、そのような評価を実現するためには、プログラム特徴とそれを説明するための言語表現の結びつきを知識記述した資源を構築し、その資源を前提に教育教材・教育方法を開発・研究することが必要と考えた。

2. 研究の目的

プログラム理解・説明の上で必要となる語や概念を整理したプログラム説明概念辞書の構築を行い、それに基づいて、プログラム理解の分かりやすい説明の言語構造やその特質を明らかにする。また、その知見を利用し、大学生および企業新入社員を対象としたプログラムの理解力・説明力を向上させる日本語による教材および教育法の研究開発を行う。

3. 研究の方法

本研究では、プログラム教育を視野に、プログラムの相互理解や説明の上で必要となる概念を言語表現の観点から整理したプログラム説明概念辞書の構築を行いつつ研究を進めた。

また、プログラム説明概念辞書と相関を持つプログラム学習者のデータは、研究活動の中で蓄積・分析することが必須となるため、プログラム教育の実践と並行させながら研

究を進めた。具体的には、プログラム要求から単にプログラムを作成する既存のプログラム教育法ではなく、抽象的なプログラム理解を言葉により説明するプレゼンテーション形式やペアワーク形式等の説明活動を研究開発しつつ、学習者がプログラム説明に用いる言語表現を蓄積し、その特質を分析しつつ研究を進めた。

プログラム説明概念は、プログラムを説明するために使用する言語表現に表出する語と語の関係を、ネットワーク構造を用いて表現することを想定する。その表現を前提に、教育上重要と考えられるプログラムの特徴的構造と、それに対応する概念表現上の部分グラフとを結びつけ、プログラムを説明する概念とその言語表現の対応を辞書として構築する。そのような辞書を整備・構築する上では、プログラムの構造的特徴や、それを適切に説明するための言語表現を対応付ける基準が必要となる。その基準を客観化し、整備・構築を支援するため、プログラムソースコードに対するリポジトリマイニングと、プログラム説明文章に対するテキストマイニングの2方向からのマイニング手法を研究開発する形で行なった。また、それぞれのマイニング手法の研究開発には、機械学習の技術を積極的に導入した。

4. 研究成果

(1) プログラム初学者への教育実践および学習者の言語表現収集として、研究初年度の2015年度より e-learning システム上で動作するプログラム学習教材を試作し、教育実践を行った。具体的な教材内容としては、プログラミングの基本構造を、初学者に自然言語で説明記述させるものであり、当初システムでは教員の経験による採点とフィードバックを行った。

この教育実践により、プログラム初学者がプログラムに用いる説明表現を収集することができ、その特質を分析した研究を関連する学会のワークショップで紹介し、研究の方向性やプログラム課題について広く意見を求めた。なお、このような学習者に対する実験的な教育方法の題材となるプログラムの基本構造については、複数の教科書のテキストデータおよびプログラムを収集・整理しつつ行った。

2016年度以降は、本節(3)(4)で記すようなマイニング手法の開発・研究を援用したプログラム説明概念辞書の整備・構築を通して、プログラムの部分要素を自然言語で説明することの適切さの評価及び、学習者へ理解受容性のフィードバックを客観的に行うことを基本に、教材・教育法の開発を行った。それあらの開発・研究成果については、教育系の学会の研究会において発表し、プログラム教育に興味を持つ研究者・実践家と知見を共有した。

以上の成果・知見は、既に e-learning 上

の選択形式の学習コンテンツ開発に活用できているが、今後この成果は、より自由な回答形式の自動評価や対話的アドバイス生成の研究開発に発展するものと考えている。

(2) プログラム初学者に向けた教育方法・教材開発に関する研究成果は上記(1)にまとめた。そのような初学者向けの試みと並行させる形で、プログラム初学者が中級へと段階的に学習・技能を高度に進展させていくための、中級学習者を主眼に据えた教育方法・教材開発を行った。

具体的には、産学連携色の強い関連学会の研究会で広く意見交換を行い、企業の情報システム開発経験者の意見を参考に、比較的長いプログラムを学習者が分析し、プログラムが実現している機能を、プログラムの基本的な部分構造や機能に切り分け、それらを基本的に他人にプログラムの内容を説明する課題を使った教育実践を実施した。そのような実践では、ペアワーク/グループワーク、プレゼンテーションや仕様書レビューといったコミュニケーション形式を取り入れた方法を開発しながら実践した。この教育実践で得られた学習者の説明記述も、プログラム説明概念を整理する上でのデータとして活用した。また、このような中級課題を題材とした学習コンテンツを e-learning システム上に試作した上で教育実践し、今後、中級以上のプログラム教育に e-learning を導入していく上での課題を整理できた。

以上の知見は、学習者の習熟度合いに応じた段階的な教材の導入・構成法とその指導に関する観点から、査読付きの誌上論文として発表した。

また、以上のような教育方法・教材開発に関わる研究成果は、企業での開発文書教育や評価に応用することが期待できる。補助期間最終年度の 2017 年度に、主に企業においてそのような興味をもつ実務家に対して、本研究の成果を共有・応用を体験するワークショップを企画・実施した。この研究成果は、今後、企業内でのシステム開発文書教育に発展するものと考えている。

(3) プログラム説明概念辞書の整理・構築を行う上で、プログラムソースコードからプログラムの特徴となる構造を抽出・整理するマイニング手法を開発した。

当初計画では、プログラム説明概念の整理はマイニング手法を援用しつつも、プログラム開発や教育の熟練者の専門知識を聴取した上で、人手によって整理することが中心的になると想定していた。補助期間開始時の 2015 年度は複数のプログラム教科書から代表的なプログラム構造を抽出する作業から開始した。しかし、本研究の企画・着想時の 2014 年前後から自然言語処理の分野で、単語の意味・知識表現に分散表現が盛んに用いられるようになり、当該分野での主な潮流とな

ると考えたため、本研究でも、それらの潮流と整合性を持つよう、次の(4)で述べる言語表現の知識記述とともに、プログラム説明概念辞書の整備・構築に計算機による二次利用が可能な知識表現を採用した。

このことにより、当初計画よりも、自然言語によるプログラム理解の説明をより客観的に分析することが可能となり、教科書等に記述されている初学者向けプログラムだけではなく、実際に実用に用いられているプログラムを分析対象として視野に入れられるようになった。

そのような視野の拡大として具体的には、インターネット上に公開されているプログラム例の蓄積(リポジトリ)に対して、良く使われ、言語表現の説明対象としても対応関係を持つようなプログラムの構造的特徴を抽出する手法の研究・開発を行った。

このような研究・開発の第一段階の成果としては、教科書に掲載されているような初学者向けのプログラムのアルゴリズムの特定や、プログラム間の類似性を算出することに貢献する中間表現を提案し、それによる特徴表現の有益性を示すことができた。この成果は、複数の学会の研究会において発表を行い、関連研究者からの意見をj得ることができた。

さらに研究・開発の第二段階の成果として、プログラム構造の特徴抽出の自動化を進めるため、木構造で表現された大量のプログラムから、特徴的なプログラム構造を自動的にマイニングする手法を提案し、その効果を実証した。具体的には、提案手法を用いて、実用向けの大規模プログラムのリポジトリから抽出した構造的特徴が、初学者向けの多様なプログラム記述を適切に分類することを確認した。この研究成果は、国内学会の研究会で発表するだけではなく、2つの国際会議で、査読付き発表として採択され、広く研究成果を公開することができた。

現在のところ、上記提案手法で得られたプログラム構造の特徴を、プログラム説明概念に結び付ける作業は、人手による整理が介在する半自動的処理で実現している段階にある。具体的には、リポジトリマイニングの結果が、本研究の研究目的である、特定の教育目的におけるプログラム判別や評価に限定して成果を示すに留まっている。しかし、この成果の今後の展開として、教育分野の応用だけではなく、自然言語からの自動プログラム生成や、プログラムの自動説明等のソフトウェア工学分野での応用や研究開発が期待できる。

(4) 本研究では、特定のプログラムを説明する際に用いられる言語表現中の語を扱うために分散表現を導入した。研究開始当初は、プログラム説明概念の整理は、初学者向け教科書や教材を分析することから開始した。分散表現の導入は、当初目的の上では、教科書間の表現の揺れや、学習者が使用する言語表

現との違いを客観的に明らかにするために役立ち、既に本節(1)(2)で記した研究を支援することに役立った。

しかし、研究が進展する中で、既存の分散表現取得ツールでは、本研究が目的とするようなプログラム説明概念を表現する上ではいくつかの課題があることが明らかになってきた。また、上記(3)で示したような、プログラムの大量事例から未知のプログラム構造の特徴をマイニングすることと連携し、計算機上で処理することが容易な表現方法として改良する必要性を認識した。

そのような認識から、一般的な文章集合から獲得された語の分散表現を、少量の専門分野の説明文書の内容に応じて、分散表現を再訓練する研究を行った。具体的には、大量の文章が入手可能なWikipedia記事を使って一般的な分散表現を得ておき、それに対して、特定目的で記述された文書群を使って再訓練し、特定目的の説明的な特徴や概念構造を反映した語間関係を算出可能な表現に改良することができた。

研究は、複数の国内学会のシンポジウムやワークショップ、研究会にて発表し、計算知能学や機械学習を専門とする関連研究者と意見交換しながら研究を進めた。成果としては、プログラム構造の特徴と対応付け可能な言語表現をより正確に特定することに寄与する語間関係を計算可能となった。また、同時に、プログラムの理解や説明の語間の関係性が、一般的な概念とどう相違するかをより客観的に分析可能となった。

本研究で達成確認した成果は、上述の通り、教育的な文脈での応用に限定されたものである。しかし、本研究での提案手法は、対象をより広げていくことにより、文章の説明目的の類型化と、より高度な説明談話の構造に関わる自然言語処理の発展に寄与すると考えている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計1件)

竹内和広, 越前拓真: プログラム学習における専門的概念導入に用いられる表現の調査, 大阪電気大学研究論集人間科学編, 査読有, 第20号, pp.93-105, 2018.

[学会発表](計10件)

Yoshiki Mashima, Takuya Okada, Sachio Hirokawa, Kazuhiro Takeuchi: Predicting Purpose of Program from Superficial Structure, The First International Symposium on Artificial Intelligence for ASEAN Development (ASEAN-AI2018), pp.45-53, 2018.
大賀賢志, 竹内和広: コンピュータ専

門用語に対する分散表現の獲得, 日本知能情報ファジィ学会第28回ソフトサイエンス・ワークショップ, pp.23-26, 2018.

竹内和広, 間嶋義喜: プログラム学習における説明文の使用語彙に関わる調査, 情報処理学会 研究報告コンピュータと教育 (CE), Vol.2018-CE-143, No.25, 7pages, 2018.

Yoshiki Mashima and Kazuhiro Takeuchi: Vector Representation of Abstract Program Tree for Assessing Algorithm Variety for the Same Purpose, Springer International Publishing AG, part of Springer Nature 2018 V.-N. Huynh et al. (Eds.): IUKM 2018, LNCS 10758, pp.126-137, 2018.
https://doi.org/10.1007/978-3-319-75429-1_11

間嶋義喜, 岡田拓也, 竹内和広: アルゴリズム推定に向けたAutoencoderに基づくプログラムのベクトル表現, 情報処理学会 研究報告コンピュータと教育(CE), Vol. 2017-CE-142, No.17, 7 pages, 2017.

間嶋義喜, 竹内和広: 同一目的アルゴリズムの多様性分析に向けたプログラムのベクトル表現, 第30回バイオメディカル・ファジィ・システム学会年次大会講演論文集 (BMFSA2017), pp.96-101, 2017.

清水豪, 竹内和広: 語の分散表現を利用した分野依存性の高い固有表現抽出, 第44回知能システムシンポジウム, SY0004-17-B5-1, 4 pages, 2017.

間嶋義喜, 大賀賢志, 竹内和広: 機能に基づくプログラム構造比較に向けた類似度の策定, 情報処理学会研究報告, Vol. 2017-CE-138, No.12, 6 pages, 2017.

大賀賢志, 間嶋義喜, 岡田拓也, 竹内和広: ソースコード内の典型的部分構造に対する説明の調査, 電子情報通信学会技術研究報告「思考と言語」, TL2016-36, pp.59-62, 2016.

竹内和広, 宮本拓弥, 山下大貴: 情報教育のためのプログラム説明における語使用の調査, 第26回ソフトサイエンス・ワークショップ, pp.22-23, 2016.

[図書](計0件)

[産業財産権]

出願状況(計0件)

名称:
発明者:
権利者:
種類:

番号：
出願年月日：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

ホームページ等

<http://takeilab.org/Syseng>

6. 研究組織

(1) 研究代表者

竹内 和広 (TAKEUCHI, Kazuhiro)

大阪電気通信大学・情報通信工学部・教授

研究者番号：20440951

(2) 研究分担者

(3) 連携研究者

兼宗 進 (KANEMUNE, Susumu)

大阪電気通信大学・工学部・教授

研究者番号：00377045

(4) 研究協力者