

「ソフトウェア開発データの分析と応用」セッション紹介

阿 萬 裕 久^{†1} 伊 原 彰 紀^{†2}

本稿では「ソフトウェア開発データの分析と応用」セッションの紹介を行う。

Introduction of “Software Development Data Analysis and Application” Session

HIROHISA AMAN^{†1} and AKINORI IHARA^{†2}

This paper introduces “Software Development Data Analysis and Application” session.

1. はじめに

本セッションは、ソフトウェア開発データの収集、分析及び応用に関して、産学双方の立場から幅広い提案・話題提供の場となるよう、昨年度の“ウィンターワークショップ 2015・イン・宜野湾”に引き続いて開催するものである。一昨年度のウィンターワークショップでは、“ソフトウェア開発データの分析”というセッションと“リポジトリマイニングの実践”というセッションの二つであったが、発表・議論される内容やそれぞれの参加者の研究テーマが近いということから、昨年度に合併して立ち上げたセッションである。昨年9月には、同テーマのセッションを“ソフトウェアエンジニアリングシンポジウム 2015”の併設ワークショップでも開催し、昨年度のウィンターワークショップで目的として掲げていた“情報共有サイト（Wiki ページ）の構築”を開始した。今回もこれまでと同様に、各参加者の現在の取り組みについてご紹介いただき、議論を深めつつ、可能なところから情報共有サイトの充実化を図っていききたい。また、解説記事やサーベイ論文の共同執筆の可能性についても検討していきたい。

2. 論文の募集と投稿内容

本セッションでは今回、図1に示す討論テーマで論文の募集を行った。

これに対して多くの投稿をいただき、全部で13件のポジションペーパーを採録とした。組織別の内訳は、

近年、産学の双方において「ビッグデータ」が一つの重要なトピックになってきている。ソフトウェア開発も例外ではなく、ソースコードやその開発に関わる作業履歴、不具合情報といった多種多様で膨大なデータに対してマイニング技術や統計解析手法を適用し、品質の維持・向上に向けたさまざまな取り組みが研究・実践されてきている。昨今では、ソフトウェア工学と他分野（自然言語処理、知能情報処理、ゲーム理論、社会学、心理学等）との融合研究が増加しており、さらなる研究分野の発展が期待される。

本セッションは、そのような開発データの分析方法並びに開発・管理への応用方法について産学双方の立場から情報の共有と議論を行い、より優れた手法の開発と実践に向けた取り組みを支援する場としたい。具体的には、参加者の皆様から現在の取り組みについてポジションペーパーを集め、その内容について議論を行うとともに、(1) 開発データの共有、(2) 手法やツール情報の共有、(3) 研究（論文）情報の共有を目的とした情報共有サイト（Wiki）の構築を行っていききたい。

さらには、データセットや関連ツール等に関する解説の執筆といった活動も視野に入れている。これらの内容に興味のある皆様の、産学双方からの幅広い情報・話題提供を歓迎する。

図1 論文募集で掲げた討論テーマ

大学から9件（五十音順：愛媛大学1、岡山県立大学1、熊本大学1、東京学芸大学1、名古屋大学1、奈良先端科学技術大学院大学2、福岡工業大学1、早稲田大学1）、企業から4件（五十音順：NTT データ1、日本IBM2、富士フイルムソフトウェア1）であり、産学双方から多く参加していただけるセッションとなった。

ポジションペーパーの内容に基づき、各参加者にお話いただく内容を簡単に紹介しておく。

†1 愛媛大学

Ehime University

†2 奈良先端科学技術大学院大学

Nara Institute of Science and Technology

- 愛媛大学・山内氏からは、オープンソース開発における各開発者の貢献度の定量化とそれらの品質との関係について、実際のプロジェクトから収集されたデータの分析結果についてご紹介いただく。
- 岡山県立大学・天寄氏からは、他組織のプロジェクトデータを用いた cross-company 不具合予測手法を“自組織のプロジェクトデータを用いる不具合予測”に適用した場合の予測精度の比較についてお話いただく。
- 熊本大学・眞鍋氏からは、ソースパッケージ分析に関する知見の共有として、Linux ディストリビューションの一つである Fedora19 のソフトウェアパッケージを分析に用いる際に生じた諸問題とその対処についてご紹介いただく。
- 東京学芸大学・杏澤氏からは、プログラム理解における重要なサブタスクの一つである Information seeking を支援する取り組みとして、ソーシャルコーディングツール上に存在する情報を使った可視化ツールの開発についてお話いただく。
- 名古屋大学・安藤氏からは、オープンソース開発におけるパッチの採否について、ソースコードの静的解析結果に基づいた予測手法の提案と Qt プロジェクトに対するケーススタディについてお話いただく。
- 奈良先端科学技術大学院大学・南氏からは、オープンソース開発における継続的なテストコード保守の実現に向けて、プロダクトコード変更時に変更すべきテストコードを特定する手法についてご提案いただく。
- 奈良先端科学技術大学院大学・平尾氏からは、オープンソース開発におけるパッチの検証作業において、各開発者が検証依頼を引き受けるかどうかを開発者の専門性やオープン性、タスク量といった観点から予測する手法についてお話いただく。
- 福岡工業大学・戸田氏からは、リポジトリマイニングを実践するにあたって現実が必要となってくる前処理に関して、CI ツールが及ぼす影響とその除去、並びにそれらに関する諸問題についてお話いただく。
- 早稲田大学・本田氏からは、オープンソースソフトウェアの評価に関する取り組みの一つとして、バグの発生状況とソースコードの変更回数との関係分析手法、並びにそれに関するケーススタディについてお話いただく。
- NTT データ・伏田氏からは、産業界におけるデータ分析の実践報告として、NTT データにおける

開発データの収集及び分析の事例についてご紹介いただく。

- 日本 IBM・青田氏からは、現場における品質や生産性の向上に向け、代表的なコードメトリクスがかかえている諸問題とその解決についてお話いただく。
- 日本 IBM・細川氏からは、ソフトウェア欠陥の類似性や分類・粒度に関する議論、さらにはソフトウェア欠陥標本の作成・共有・移転・流通させる技術について紹介いただく。
- 富士フイルムソフトウェア・小室氏からは、欠陥摘出数に関するいくつかの数理モデルとそれらに基づいた品質予測モデルの構築、並びにそれらを実プロジェクトに適用した結果についてお話いただく。

3. おわりに

本セッションでは、前述したように産学双方から多くの参加があり、当日はソフトウェア開発・保守に関するデータの収集と分析、応用に関する幅広い話題の提供が期待される。議論を通じて、各自にとって有益な情報交換の場となるようにしていきたいと考えている。