

問題解決過程を考慮した不具合報告の 要約生成手法の拡張

60246164 田中智達

背景

| ソフトウェアプロジェクトでは日々多くの不具合報告がなされている。

| プロジェクトを運用している開発者らがバグを迅速に取り除くため、バグ管理システムが用いられている。

- どのようなバグが存在するのか、その内容について詳細に記述し管理することで円滑な対処を促進。
- さまざまなステークホルダによって多くのバグ報告がおこなわれることで、潜在的なバグをなくす。

背景

| しかし多数の不具合報告がされるようになり，かつバグの内容も詳細に書かれるため，開発者が効率的にバグの対処を試みることが難しくなっている．

| そのため，バグ票の理解を容易にし迅速に対応できるような支援が必要．



| 開発者支援を目的として不具合報告を要約する研究が行われている．

先行研究

| 先行研究では内容把握を目的とした手法と情報収集を目的とした手法がある。

| 内容把握を目的とした手法

- 不具合報告票に含まれる問題点を要約することに重点。
- 開発者が問題解決をする際に必要としている情報が含まれている保証がないという問題。

| 情報収集を目的とする手法

- 問題と要因と解決という3つの要素が含まれるようにテキストを要約するということに重点。
- 上記のテキストに含まれる品詞の分布の偏りを用いる際のベースとなるデータ量が絶対的に不足しているため、構築した手法の分類性能が十分に発揮されているとは言えない問題。

3つの要素と含む理由

問題

- 報告者が抱える問題の把握

要因

- 不具合が発生した原因の把握

解決

- 不具合の解決方法の把握

いずれも個々の開発者やユーザが抱える問題を解決するために、過去に報告・解決された不具合の情報を効率的に入手することを目的。

目的とアプローチ

目的

- 句構造重みベクトルを構築する上で必要な頻出ルールを抽出するためのデータ量を増やし、分類性能を向上させる.

アプローチ

- 不具合報告票の中から3つの要素を含むテキストを抽出し、品詞の分布を再計算する.

データセット

| 不具合報告票

- FirefoxのBugzillaから2537件, 11763文のデータを収集.
- 目視で3つの要素のある文を抽出.
 - 評価基準は先行研究とほぼ同等の精度(99.56%).

今後の流れ

1. 11763文を目視で分類
 2. クリーニングして単語ごとに分割
 - テキスト特徴量抽出のため
 3. 句構造の特徴を用いて問題文要因文解決文の特徴量を捉え、ベクトルに変換
 4. テキスト特徴量と問題文要因文解決文の特徴量を捉えた句構造重みベクトルを用いて要約を生成
- 並行して卒論を書く