

機能に基づいたアプリ検索手法の提案

上西達也

新規アプリ普及の課題

- スマートフォンなどのアプリケーション(アプリ)は大量に出回っており、新規アプリも日々開発、リリースされている
- ユーザはそこから自由に選んで端末にインストールするが、**すでに人気のアプリを採用するユーザ**もいれば、**自ら好みのアプリを探して採用するユーザ**もいる

開発者が新規アプリを普及させたいときは、まず **自らアプリを探すユーザ** にアプリを見つけてもらう必要がある

自らアプリを探すユーザーの特徴

- 流行に敏感で、新しければとりあえず採用してみる姿勢、冒険心がある^[1]
- 自らの利益を考慮しながら採用する(リスク許容度が低い)人も多い^[1]



アプリ購入前にその機能を把握し、求めている機能を持ったアプリを選びたい
欲求があると考えられる

[1] イノベーション普及学, Everett M. Rogers

現在のアプリ検索システムの問題点

1. タイトルや説明文などの記述内容によって順位が変わる
 - 開発者がそのアプリの特徴を正確に記述できるとは限らない
 - アプリストア最適化 (ASO) などの特殊な知識が必要
2. すでに人気のあるアプリを上位に表示する傾向がある
 - ダウンロード数が少ない新規アプリなどはたとえ優れていてもユーザーの目に触れにくい

研究の目的

- 自らアプリを探すユーザは人気よりも**求めている機能があるか**（利益があるか）に着目する
- **機能に基づいたアプリ検索システム**があれば、自らアプリを探すユーザが求めている機能を持ったアプリを見つけやすくなる



本研究では、機能に基づいたアプリ検索手法を提案する

対象とするユースケース

既知のアプリと機能面で類似するアプリを検索する場合

- 欲しい機能があるもの
- いない機能がないもの

アプリ解析における課題

- アプリの多くはソースコードが公開されていない
- 従来からソフトウェア検索手法に関する研究はあったが、ソースコードが手に入ることが前提となっている
- 既存手法は使えない
- 実行可能ファイルなどから抽出できる情報のみを使った手法が必要



本研究では、Stigmataというツールを応用する

Stigmata

- ソフトウェアの盗用を発見、証明するための研究にて開発されたツール
- Javaのクラスファイルから、そのソフトウェア固有の特徴(ソフトウェアバースマーク)を抽出できる
- 例えば、JARファイルから標準APIメソッド呼び出し列などを抽出できる
 - 多くの難読化ツールが標準APIメソッド呼び出しまでは難読化しないことを利用している

提案手法の概要

1. Stigmataを用いてアプリ解析
2. アプリをBag of Words共起行列で表現
3. 潜在的意味インデキシング(LSI)を適用？
4. 既知のアプリ(検索の入力)の各クラスと同じ機能を持つと推定されるクラスをより多く持つアプリを提示？

1. アプリ解析

一般的なアプリ開発では、標準APIを組み合わせて目的の機能を実装する



Stigmataを用いて、アプリの標準APIメソッド呼び出し列の集合を抽出する

(ここに出力結果の表を例示)

2. アプリの行列表現

各クラス内の標準APIメソッド呼び出しの組み合わせの類似性(共起関係の類似性)は機能と関連があると考えられる



Stigmataの出力から、クラスを文書、標準API呼び出しを単語に対応させて共起行列(Bag of Wordsモデル)を作成して利用する

ケーススタディ

- Androidアプリを対象とする
 - 実行可能ファイルはAPKファイル(JARファイルに変換できる)

現在の状況

- 実験中
- どんな実験をすべきか
- どんなアプリを入れるべきか