Android アプリの権限要求に対する説明十分性の自動確認システムの提案

小島　智樹†　　酒井　哲也†

† 早稲田大学基幹理工学部情報理工学科　〒 169–8555 東京都新宿区大久保 3-4-1

E-mail: †frkojima512@ruri.waseda.jp, ††tetsuyasakai@acm.org

あらまし　Anrdoid 端末においてアプリケーションは端末上のセンシティブな情報にアクセスする権限をユーザーに要求する．その権限を要求する方法として，最近では Runtime Permission システムという実際に必要な機能へアクセスする際にユーザに許可を求める方式が主流である．それに付随して，アプリの開発者は使用理由に関する Rationale(理論的根拠) をダイアログで提示することができる．これによって．ユーザはより安心してアプリによるアクセスを受け入れることができる．しかしながら， Liu らの調査 [1] によればこのダイアログによって提示される説明は未だに不正確なものが多いとされている．そこで本研究ではダイアログによって提示される説明の不足について自動判定を行うシステム，Permission Rationale Checker (PRC) の構築を行う．その際に，従来研究で用いられてきた分類対象の文章だけでなく，アプリケーションの説明文 (description) を利用することで精度の向上を試みた．その結果，description を利用した word2vec の潜在表現の学習は分類における F1 値の改善に寄与することが判明した．また，description に対して適切な前処理を施すことにより更なる F1 値の改善が見られた．ただし，各システムに対して正解率について統計的有意差は確認されなかった．

キーワード　Google Play,Application,Permission

1. はじめに

Permission とは，本論文では Android OS の一部の機能や情報にアプリがアクセスするための許可のことを表す．機能の例として，位置情報やマイク機能といったものが挙げられる．これらはどれもユーザーにとってセンシティブな情報である．Bonné らの研究 [2] で行った実験では調査対象の 44%がアプリによる権限の要求に不快感を示した．そのため，ユーザーが安心してアプリを使用するためにはアプリの開発者は十分な説明を行うことが望まれる．そこで用いられるのが Rationale Text である．Rationale とは日本語で理論的根拠のことであり，本研究においては Rationale Text(以後 RT と記述する) を「権限をユーザーに求める際に行われる説明」という意味で使用する．アプリが RT を提示することによって，アプリに権限を与えることへのユーザーの安心感を高め，承認率の向上を期待することができる．

RT は主に 2 種類の方法で提示される．

1 つ目の方法としては Google Play などのアプリストア上での説明がある．図 1 は Google Play で公開されているアプリ，「Yahoo!乗換案内」[3] の説明文の一部である．ユーザーはこのような説明により，安心感を持ってアプリに権限を与えるかの決定をすることができる．

別の方法として，バージョン 6.0 以降の permission の機構である Runtime permission を拡張したものを使う場合がある．図 2(a) は Google Play で公開されているアプリ「Facebook」を起動した際に表示される画面である．図 2(a) が開発者が RT を表示するため，独自に設定したダイアログである．その後，

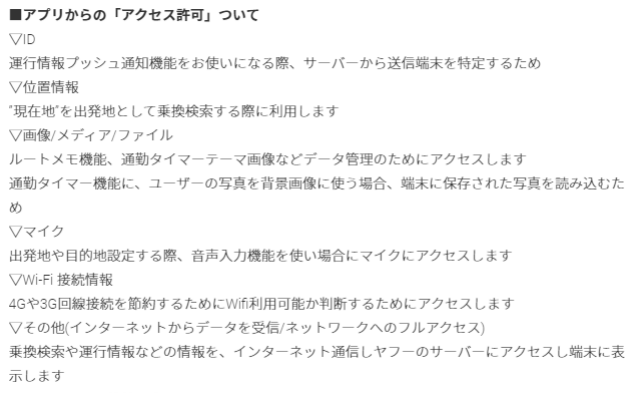

図 1: Yahoo!乗換案内の説明文

(a) 開発者が独自に設定したダイアログ　　(b) システムのダイアログ

図 2: Facebook 内で表示先ダイアログ

図 2(b) のようなシステムのダイアログが出現し，実際にその権限を許可/不許可を選択する．この方法では実際に権限を必

要とする際に理由を提示する．これはコンテキストを踏まえた上での説明であるためユーザーはより必要な理由を理解でき，それにより承認率の向上にもつながる．

しかしながら，Liu らの研究 [1] によればこのように RT が表示されている場合でも文言が適切でない場合がある．適切な RT の設定による承諾率の変化は Tan らの研究 [4] によって示されている．この研究では使用する権限の種類とその理由を十分に明記した場合は単に使用するということを提示した場合に比べ，承諾率が最大で約 20%向上したという結果が示されている．このように，適切な RT の設定はユーザーにとっては納得感を持った権限の受け入れの助けになり，それによって開発者は自身が開発したアプリの使用に繋がるため重要である．

本研究ではある RT が適切な説明であるかの自動判定機構，Permission Rationale Checker (PRC) の提案を行う．PRC の学習には RT のみを用いる方法と分類対象のアプリケーションの説明文 (以後と呼ぶ Description) を補助データとして使用する方法で学習を行う．実験の結果，統計的有意差は観られなかったものの，Description の補助的なデータとして利用することで F1 値の改善が見られた．

以下に本論文の構成を示す．2 章において，本研究に関連した Android のセキュリティやユーザーのプライバシーへの意識，Permission システムなどに関する研究を紹介する．3 章において，PRC に関する構成や関連技術について述べる．4 章において，それら機構が上手く機能しているかについて実験を行った結果を記す．5 章において，今後の課題について記す．

2. 関 連 研 究

2.1 ユーザーのアプリケーションのセキュリティに対する意識

アプリケーションのユーザーのセキュリティに対する意識を調査した研究として Bonné らの研究 [2] や Golbeck らの研究 [5] が挙げられる．Bonné らの研究では Runtime permission が権限の承諾にどのような影響を与えるかを様々な権限に対して調査している．また Golbeck らの研究では Facebook を例にユーザーはどのような情報へのアクセスについての理解しているのか，またそれに対してどのような懸念を抱いているのかなど，セキュリティへの意識について調査している．

2.2 Android アプリに関するテキストの分類

アプリの Rationale に関する先行研究はまだ多くない．そのため，ここではアプリと自然言語処理に関する論文について述べる．

Nayebi らの論文 [6] ではアプリの Review の分類を行っている．この論文中では Review の分類を行う際に追加のデータとして Review に付随するレーティングを利用している．

Review を利用した Android アプリの分類に関しては Gómez らの研究 [7] がある．これは Latent Dirichlet Allocation (LDA) による教師なし学習を利用し，Review からバグの発生しやすい権限のパターンについて解析する研究である．

また，Jha らの研究 [8] では Bag Of Frame という手法が使われている．これはアプリの Review を構成する単語を抽象度の高い Frame という単位に変換することで，表現の揺れに強く過学習を抑制した特徴を作成し高い分類精度を実現している．この研究にあるように，アンドロイドアプリに付随する文は表現が多様である．そのため，単語という具体的な特徴で捉えるよりも，より抽象度の高い特徴で捉える方が正確に文意を捉えることが可能になると考えられる．それを行うため，本研究では word2vec(w2v) を用いた．

2.3 権限の付与の提案に関する研究

ユーザの権限に関する研究として，Liu らの研究 [9] がある．これはユーザーの権限の許可/不許可に関するログデータを利用し，ユーザー意思を推測することで複数権限の On/Off に関する提案を行い，ユーザーの決定の手間を減らすシステムである．これによって意思決定の回数を減らし，より考えて権限の付与を行うことが可能になる．

2.4 Rationale の有無の自動判定について

アプリ以外に対する Rationale の有無について行われている研究として，Kurtanovic らの研究 [10] や Alkadhi らの研究 [11] がある．

Kurtanovic らの研究の対象はユーザによるソフトウェアの Review である．それらを分析し，どのような理由でソフトウェアが評価されているか，改善を求められているかなどについて調査を行っている．

Alkadhi らの研究においてはソフトウェア開発者のバグレポートについて同様の行為を行っている．

3. 提 案 手 法

Permission Rationale Checker (PRC) は

（1） テキストの前処理

（2） 文のベクトル化

（3） 学習と分類

の 3 ステップで構成される．

テキストの前処理について，description と RT の共通の前処理に関しては，以下を実施した．

- stemming
- レンマ化
- ストップワードの除去
- 文字の小文字化

それに加えて description の前処理に関しては追加で以下を実施した．

- URL の除去
- 正規表現を用いた記号の除去
- 形容詞の除去

形容詞除去は本研究独自の試みである．description に含まれる形容詞は主にアプリケーションの特徴の強調やユーザーへの宣伝のために用いられる．そのため，word2vec の学習の際にノイズになると判断して除去を行った．

記号の除去に関しては以下の通りに行った

- 記号の後が大文字であれば，それは文を区切っている記号と判断し. に置換
- 記号の後が小文字であれば，単語を区切っている記号と

判断し” ” (スペース) と置換

文のベクトル化については次の章で述べる．ベクトル化した後，それらを学習データとして RnadomForestClassifier を学習，分類を行う．

4. 評 価 実 験

以下で PRC の評価実験の概要について述べる．

4.1 データセット

4.1.1 Rationale Text

RT のデータとして Liu ら [1] が論文内で使用している，公開データ [12] を利用する．このデータセットは様々な権限に対する RT のデータである．これらのデータがユーザーが権限を承諾する際に十分な説明になっているか，自動で判定するシステムを構築する．

学習データ，及び教師データとするために，まず最初に英語以外のデータを除外する．次にデータに対してラベリングを行う．正例のラベルは以下の基準を全て満たすものに付けられ，残りを負例とした

（1） どんな機能のために権限を利用するか明示されている

（2） その機能が自然言語で具体的に明記されているか

（3） その機能が権限に適切か

（4） 対象となる権限 (今回はマイク) のための説明であるか

正例の具体例を以下に示す．

- to use voice search , allow android tv remote permission to record audio .
- voice based typing cannot work without audio recording and storage permissions
- without microphone and location permission the cruise finder voice search will not work .

負例の具体例を以下に示す．

- to use the application , the microphone and storage permission is required . (1 つ目の基準を満たしていない)
- uh oh ! we can t access your microphone ! (1 つ目および 2 つ目の基準を満たしていない)
- triller would like to use your camera to record your videos (4 つ目の基準を満たしていない．カメラへの権限の要求の理由は説明されているが，マイクについての言及が無い)
- you must accept the microphone permission to be able to use this feature . (2 つ目の基準を満たしていない)

先述の基準に沿って分類された各データは以下の通りの個数である．

表 1: RT データの内訳

	カウント	平均単語数
正例	629	14.37
負例	987	12.19
正+負	1616	13.03

4.1.2 description データ

今回の実験では分類対象の RT のデータ以外に分類対象のアプリの description のデータを用いた．これを用いることによって w2v の学習を行うことで Android アプリ特有のドメイン知識を学習させることが目的である．以後単に description と呼ぶ．また，そのために分類対象のアプリ以外の説明文も追加のデータとして使用した．以降これを extra-description と呼ぶ．内訳は以下のとおりである．

表 2: description データの内訳

	アプリ数	文の平均長
description	1616	290.56
extra-description	52718	217.38

4.2 分類器とデータの分割

評価は RT を 3:1 で学習データとテストデータに分割し，それを用いた 2 値分類の Precision，Recall，F1 値で行う．各指標は TP を「正例のものを精霊と予測した数」，FP を「負例のものを正例と予測した数」，FN を「正例のものを負例と予測した数」としたとき，以下の式で表される

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1\text{ 値} = \frac{2Recall \cdot Precision}{Recall + Precision} \tag{3}$$

また，分類器には RandomForestClassifier を用いた．

4.3 ベクトル化

4.3.1 ベースライン

本研究では複数の方法で RT をベクトル化し，その際の Precision，Recall，F1 値を比較する．

まず，bag of words と 2-gram による RT のベクトル化をベースラインの一つとする．これを用いたモデルを model1 と呼ぶ．

次に Google が公開している word2vec のモデル [**?**] を用いた方法を 2 つ目のベースラインとして用いる．この word2vec のモデルは Google ニュースデータセットの一部で学習された 300 次元，300 万単語に関するモデルである．これを用いて，RT の各単語をベクトル化した後にそのベクトルの平均を取ることで文章のベクトル化を行う．これを用いたモデルを model2 と呼ぶ．

4.3.2 提 案 手 法

次に，アプリのデータを用いて学習させた word2vec のモデルについて説明をする．これは学習に用いるデータによって以下の 3 種類に分けられる

（1） RT のみを学習データとして学習させた w2v を用いる方法

（2） RT と description を学習データとして学習させた w2v を用いる方法

（3） RT と description，更に extra-description を学習データとして学習させた w2v を用いる方法

それぞれの方法で w2v を学習させ，RT の各単語をベクトル化しそれの平均を文章ベクトルとする．各方法でベクトル化したものをそれぞれ model3，model4，model5 とする．

4.4 形容詞の除去における分類性能の変化

アプリケーションの description において 形容詞が多い文はアプリケーションの宣伝，ユーザーへのアピールといった RT の分類に対して効果の小さいとされる文が，それは w2v の学習の際にノイズとなりうる．そのため，description から形容詞を除去し w2v の学習を行った．model4 の学習の際に形容詞除去を行ったものを model6,model5 の学習の際に形容詞除去を行ったものを model7 とした．

4.5 分類結果

結果を表 3 に示す.

表 3: 分類結果

	Precision	Recall	F1score
model1	**0.782**	0.477	0.592
model2	0.639	0.609	0.624
model3	0.594	0.617	0.605
model4	0.614	0.633	0.623
model5	0.622	0.578	0.599
model6	0.621	0.641	0.631
model7	0.634	**0.664**	**0.649**

実験の結果．形容詞除去と extra-description を用いたモデルである model7 が最も高い F1 値を達成した．なお，各システムに対して正解率について Tukey の HSD 検定を行ったが有意差は確認できなかった．

5. 結論と今後の課題

本研究では Permission Rationale Checker (PRC) の提案と作成を行った．これは Rationale Text がユーザーにとって権限を与えるための一助となりうるかを判定するものである．

本研究における分類対象である RT は概して短く，さらに先行研究のようにストアにおけるユーザーからの評価といったメタデータを直接用いることは難しい．そのため，今回はその足りないデータを補うために説明文を利用し，w2v の学習の際のデータ量を増やすという形で利用した．

分類器としては RandomForest を使用し，文章のベクトル化としては bow や複数の方法で学習させた w2v を用いた．

その結果，RT と extra-description を用いて w2v を学習させたものが最高の F1 値を記録した．比較対象の Google ニュースデータセットを用いて学習させた w2v よりも高い精度を記録しており，このことからドメインに特化した文章による学習が有効に作用していると考えられる．ただし，正解率について統計的有意差は今回確認できなかったため，更なるデータによる検証が必要である．

今回の実験ではマイクの権限に関する RT を使用したが，これは比較的ユーザにとって理解しやすい権限でありセンシティブではあるものの，なぜ使われるか想像しやすい場合が多い．そのため，今後の実験においてはより表面的に何に用いられるかわかりにくい権限についても分析を行う．

また，RT はユーザーの目に直接触れるものである．そのため，今後はユーザーインタビューを行い，定性的な結果を得る必要がある

文　献

[1] Xueqing Liu, Yue Leng, Wei Yang, Wenyu Wang, Chengxiang Zhai, and Tao Xie. A large-scale empirical study on android runtime-permission rationale messages. In Caitlin Kelleher, Gregor Engels, Joao Paulo Fernandes, Jacome Cunha, and Jorge Mendes, editors, *Proceedings - 2018 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC 2018*, Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC, pp. 137–146. IEEE Computer Society, 10 2018.

[2] Bram Bonné, Sai Teja Peddinti, Igor Bilogrevic, and Nina Taft. Exploring decision making with android's runtime permission dialogs using in-context surveys. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pp. 195–210, Santa Clara, CA, July 2017. USENIX Association.

[3] Yahoo!乗換案内　無料の時刻表、運行情報、乗り換え検索.

[4] Joshua Tan, Khanh Nguyen, Michael Theodorides, Heidi Negrón-Arroyo, Christopher Thompson, Serge Egelman, and David Wagner. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 91–100, New York, NY, USA, 2014. Association for Computing Machinery.

[5] Jennifer Golbeck and Matthew Mauriello. User Perception of Facebook App Data Access: A Comparison of Methods and Privacy Concerns. *Future Internet*, Vol. 8, No. 4, p. 9, mar 2016.

[6] Necmiye Genc-Nayebi and Alain Abran. A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, Vol. 125, pp. 207–219, mar 2017.

[7] María Gómez, Romain Rouvoy, Martin Monperrus, and Lionel Seinturier. A Recommender System of Buggy App Checkers for App Store Moderators. In *Proceedings - 2nd ACM International Conference on Mobile Software Engineering and Systems, MOBILESoft 2015*, pp. 1–11. Institute of Electrical and Electronics Engineers Inc., sep 2015.

[8] Nishant Jha and Anas Mahmoud. Mining user requirements from application store reviews using frame semantics. In *International working conference on requirements engineering: Foundation for software quality*, pp. 273–287. Springer, 2017.

[9] Bin Liu, Jialiu Lin, and Norman Sadeh. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help? In *Proceedings of the 23rd international conference on World wide web*, pp. 201–212. ACM, 2014.

[10] Zijad Kurtanovic and Walid Maalej. Mining User Rationale from Software Reviews. In *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference, RE 2017*, pp. 61–70. Institute of Electrical and Electronics Engineers Inc., 2017.

[11] Rana Alkadhi, Teodora Lata, Emitza Guzmany, and Bernd Bruegge. Rationale in Development Chat Messages: An Exploratory Study. In *IEEE International Working Conference on Mining Software Repositories*, Vol. 0, pp. 436–446. IEEE Computer Society, 2017.

[12] Runtime Permission Project. https://sites.google.com/view/runtimepermissionproject/home?authuser=0.