Androidアプリの権限に対するユーザーへの説明の補完

小島　智樹†　　酒井　哲也†

† 早稲田大学基幹理工学研究科情報理工・情報通信専攻 〒 1698555 東京都新宿区大久保 341
E-mail: †frkojima512@ruri.waseda.jp, ††tetsuyasakai@acm.org

あらまし　Anrdoid 端末においてユーザがアプリケーションをダウンロードするとき，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要求する．権限はときに悪用されるため，ユーザーはアプリケーションがなぜ権限を要求するかわからず，不安を覚えることがある．本論文では既存のアプリケーションの情報を利用し，権限の説明を補完するための手法について提案する．

キーワード　Android, 情報検索, 自然言語処理,UX

1 導　　入

Android は Google が開発しているスマートフォン向けのオペレーティングシステムである．2017 年，アクティブな端末数は全世界で 20 億台を超える [1]．公式のアプリケーションストアである Google Play には 300 万以上のアプリケーション [2] が有料，または無料で公開されている．アプリケーションの総ダウンロード回数は 2017 年の 4 半期のみで 190 億回を超える [3]．

ユーザがアプリケーションをダウンロードする際に，アプリケーションがデバイスのセンシティブなデータや機能を使用する権限を要求することがある．例として，ユーザの位置情報の取得への権限などが挙げられる．アプリケーションは通常，権限をユーザーにとって有益な機能の提供のために使う．しかし一方で，悪用することでユーザに害をなす事態が発生している．具体的な事例として，カメラ機能の権限を不正に利用することで盗撮を行う [4] といったものが報告されている．

このように権限は悪用されることがあるため，ユーザーは権限の要求理由がわからないときに不安を感じる．この不安を払拭するために，本来であればアプリケーションの説明文において権限を要求する理由を述べるべきである．例えば，Yahoo!乗換案内 [5] のアプリケーションの説明文を見ると各権限について要求する詳細な理由が記述されている．しかしながら，多くのアプリケーションにおいて，説明文中には十分な説明が無く，ときにユーザの不安を引き起こす．

また，昨今では Facebook などのアプリケーションにおける情報取得に対してユーザーからの不安の声が出ており，たとえ正当な要求であっても，ユーザーの安心な利用のためには権限を要求することへの説明が必要であると考えられる．

本研究では，説明が不足しているアプリケーションに対して，権限の説明を補完することで，ユーザの不安を和らげる手法について提案する．

類似の先行研究として Liu らが考案した CLAP というフレームワークがある [6]．これは英語のアプリケーションに対して既存のアプリケーションの説明文をヒントとして提示して，開発者が適切に権限に対する説明文を作成することを助ける．しかし，このフレームワークでは，説明文の作成は開発者の自主性に委ねられる．そのため説明文の品質に差がでたり，ときには作成されないと言った事態が考えられる．また，先行研究において説明文作成のヒントの提示は，アプリケーション単位であって権限単位ではない．そのため複数の権限を要求するアプリに対しても，ヒントは一括して与えられる．また，先行研究で用いられているのは英語のデータであるが，本研究ではデータ量が少ない日本語のデータを扱う．

本研究では以下の 2 点に関して提案を行う．

- 権限ごとに自動で説明を与えるフレームワークの提案.
- 日本語のデータに対して，より適切な説明を行うための英語データの使用の提案.

この章以降の本論文の構成は以下の通りである．2 章ではアンドロイドアプリに関する先行研究について述べる．特に本研究が大きく影響を受けた Liu らの研究について重点的に述べる．3 章では権限に対して説明を行うフレームワークについての提案を行う．4 章では英語のデータを日本語のアプリケーションの権限説明に用いる手法について提案する．5 章では 3，4 章で提案した手法の性能を評価する．6 章では 5 章で得られた結果に対して考察を行う．7 章で本研究の結論を述べる．

2 従 来 研 究

2.1 権限の説明に関連した従来研究

権限の説明に関連した従来研究には Liu らが考案した CLAP というフレームワークがある [6]．これは入力としてアプリケーション A の説明文，カテゴリ，権限，タイトルを使用し，出力として A に対して権限についての説明のヒントを開発者に与えるものである．大まかな流れとして以下の手順で行われる．

（1） *A* の前述の要素を利用して類似のアプリケーション *AS* をいくつか，既存のアプリケーションの中から探す.

（2） *AS* の各説明文を動詞句単位で分割.

（3） 分割に対してスコアリングを行い，その中の上位のいくつかを説明文を作成する際のヒントとして開発者に提示する.

手順の中の (1) において，説明文の類似度の計算は BM25 という重み付けの方法 [7] で文章のベクトル化を行った後，その

コサイン類似度を用いて計算される．しかし，この方法は単語ベースのベクトル化である．そのため意味的な類似性は考慮されない．例えば

- 写真を撮って共有する
- 撮影した画像をシェアする

といった 2 文があったとき，人手では 2 文が概ね同じ意味の文と認識できる．しかし，単純な単語ベースの比較では単語の一致がないため，全く異なった文と判断される．このような理由で先行研究の方法では，有用な情報が失われることが考えられる．本研究では Word2vec を用いた別の計算方法を用いることで，意味情報も考慮した類似度の判定を行う．

また，Liu らの先行研究ではアプリとの類似性を利用して文の検索を行ったが，複数の権限があるときの機能と権限との対応関係までは考慮していない．本研究では「このアプリのこの権限に対応する機能は…」といった説明を最終目標とする．

2.2 過剰な権限の検知

先行研究において，様々な方法による過剰な権限の検知が行われている．自然言語処理を用いた例としては Pandita らが提案した WHYPER [8] や Zhang らによる AutoCog [9] などがある．

Gorla らの研究 [10] ではアプリケーションの記述から分類を行い，その後 API 呼び出しを利用して異常なアプリを判別している．

その他の検知方法としては Zhang らによる API の呼び出しグラフを利用する方法 [11] や Yerima らによるベイズ分類器を用いた静的解析を利用する方法 [12] がある．

これらの研究は，不正な権限の検知という観点でユーザの安心に寄与している．

一方で，本研究においては正常な権限の説明という観点でユーザの安心させることを目指す．自然言語と言う形でユーザに提示することで，従来手法に比べて納得感が増すと考えられる．

2.3 ユーザのセキュリティ意識について

ユーザのセキュリティへの意識に関する研究には，Chin ら [13] や Lin ら [14]，Vidas ら [15] の研究がある．Chin らは複数の観点におけるユーザのセキュリティ意識についてコンピュータとスマートフォンの違い等に関して調査した．

Lin らはアプリケーションの権限へのユーザの感じ方をクラウドソーシングを利用して調査した．そしてその結果を用いて，ユーザーに追加の警告を出すことを提案した．

Vidas らはソースコードの API 呼び出しなどを解析することにより，開発者に最低限の権限を要求することを提案している．これによって，不要な権限をユーザに要求することを減らせ，不安を軽減することにつながる．

2.4 文章の類似度の研究

tf-idf や BM25 に基づく文章のベクトル化は単語の出現を基に行われる．そのため意味情報の考慮がされず，文の正確な類似度が判断できないと考えられる．そのため，意味情報を考慮した文章の類似度の計測は情報検索に不可欠である．

Mihalcea らの研究 [16] では相互情報量などを用いることで，意味情報を考慮した文の類似度の計算を行っている．

また，Kenter と de Rijke の研究 [17] では tf-idf だけでは無く，文を構成する単語を Word2vec でベクトル化したものを利用することで文同士の類似度を計測する．これによって，単語の出現だけでなく意味情報も考慮して 2 つの文の類似度を計測することができる．

2.5 複数言語を用いた情報検索

Sakai らの研究 [18] ではある言語 1 で書かれた文を検索する際のスコアリングに，他の言語 2 で書かれたデータを言語 1 に翻訳したものを用いることで検索性能の向上が見られた．

本研究においても，Sakai らの研究のように翻訳を利用した多言語データの利用を行う．それに加えて，言語非依存と考えられる関係の計算にも，多言語データの応用を行う．

3 説明付与に関する提案手法

アプリ A の権限 P への説明付与は以下の手順で行われる．

（1） 既存のアプリケーションの中で A に類似しているものを検索する

（2） 残ったアプリケーションの説明文を更に細かく区切る

（3） 区切られた説明文の中から権限 P およびアプリケーション A を利用して説明としてふさわしいもの検索する

3.1 検　　索

アプリケーション A に対して以下の要素を利用して類似アプリケーションを求める

- アプリケーションの説明文
- タイトル
- カテゴリー
- 権限

各項目に関するスコアを s_1〜s_4 として，合計スコアを

$$S = \lambda_1 s_1 + \lambda_2 s_2 + \lambda_3 s_3 + \lambda_4 s_4 \tag{1}$$

とする (各 λ は重み付けのための定数)．このスコア S を用いて類似アプリケーションを決定する．これらの要素，およびスコアリングの方法の一部は Liu らの先行研究 [6] で用いられたものである．

3.1.1 アプリケーションの説明文

説明文の類似度は Kenter と de Rijke の意味情報も考慮した方法 [17] を用いて行う．２つの説明文 s_l, s_s があるときに類似度は

$$f_{sts}(s_l, s_s) = \sum_{w \in s_l} IDF(w) \cdot \frac{sem(w,s_s)\cdot(k_1+1)}{sem(w,s_s)+k_1\cdot\left(1-b+b\cdot\frac{|s_s|}{\mathrm{avgsl}}\right)}$$

$$sem(w,s) = \max_{w' \in s} f_{sem}(w,w')$$

で求めることができる．ここの S_s, $avgsl$ はそれぞれデータに含まれる説明文の最短長および平均長，k_1 と b はパラメーター，f_{sem} は２つの単語ベクトルのコサイン類似度を取る関数である．

3.1.2 タイトル

タイトル間の類似度の計算も同じく Kenter と de Rijke の方

法を用いて行う．

3.1.3 カテゴリ

カテゴリの類似度は以下の手順で求める．

（1） 各カテゴリに属する全てのアプリの説明文を一つの文とみなす

（2） その説明文の tf-idf ベクトルを作成する

（3） 各ベクトル間のコサイン類似度を計算する

3.1.4 権限

権限の類似度は，2 つのアプリケーションの権限に関して Jaccard 係数を計算することにより取得する．

3.2 分割

文章の分割には句点を用いた分割を行った．文の分割を行うのは，Liu らの研究 [6] でも述べているように, 既存の説明文全体を用いるのはユーザにとって無駄な情報が多く含まれるからである．

3.3 権限に対応する単語の抽出

ある権限 P に対して相互情報量を用いることで，権限に特有の単語を全アプリケーションの説明文から抽出する. 対象とする単語は動詞および名詞である．これによってある権限に特有の動詞，および名詞が抽出される．相互情報量 I は以下の式で表される．

$$I(U;K) = \sum_{e_t\in\{1,0\}e_p\in\{1,0\}} P\left(U=e_t, K=e_p\right)\log_2\frac{P\left(U=e_t, K=e_p\right)}{P(U=e_t)\,P\left(K=e_p\right)} \tag{2}$$

確率変数 U は単語 t の出現を，確率変数 K を権限 p が要求されることを表す．単語 t が出現するとき $U=1$ をとり，単語 t が出現しないとき $U=0$ のときをとる．また，権限 p が要求されるとき $K=1$ を，権限 p が要求されないとき $K=0$ をとる．これによって 単語 t と権限 p の依存度がわかり, I の大きい単語の上位を選ぶことで権限に関連する単語を抽出することができる．

3.4 アプリケーションおよび権限に対応する文の決定

文の決定には，Liu らの研究のアルゴリズムを応用する．この方法は文 S の各単語 w に対して投票を行い，文で平均をとったものを文の投票とする．

先行研究において，ある単語 w への投票は

$$votes(w) = IDF(w)\times\frac{1}{K}\sum_{k=1}^{K}\frac{TextRank\left(w,D_k\right)}{\max_{w'\in V}TextRank\left(w',D_k\right)}$$

で行われた．式中の D は検索で類似していると判定されたアプリの説明文である．式中の $TextRank$ は，Rada らの方法 [19] で計算される．

本研究では，権限へ文が対応しているかも文を選択する際の重要な要素である．そこで，権限 P に関連のある単語の集合を $Pwords$ として

$$votesP(w,Pwords) = \begin{cases} IDF(w)\times\frac{1}{K}\sum_{k=1}^{K}\frac{TextRank\left(w,D_k\right)}{\max_{w'\in V}TextRank\left(w',D_k\right)} & (w\in Pwords)\\ 0 & (otherwise)\end{cases} \tag{3}$$

とすることで，権限へ関連のある単語のみを考慮するスコアリングに変更した．これを利用してある候補文 s の権限 P に関する投票は

$$votes(s) = \frac{1}{|s|}\sum_{w\in s} votesP(w,Pwords) \tag{4}$$

で計算される．この上位 5 件を権限への説明とする．

4 英語データの使用に関する提案手法

4.1 翻訳された英語の文を用いる方法

英語のデータを日本語に翻訳して用いることで，より機能の網羅性が上がり，正確な説明が可能となると考えられる．本研究では，英語の説明文を Google 翻訳 API を用いて，日本語に変換してデータを増強する．

4.2 カテゴリの類似度の測定に英語のデータを用いる方法

カテゴリ間の関連性は，カテゴリごとに権限の使われ方に特徴があるという観点から重要な情報である．よって，より正確な計算が求められる．日本語でも英語でも，カテゴリ間の類似度は概ね同じという仮定のもと，カテゴリの類似度の計算に英語のデータを用いる．これによってより多くのデータが使え，正確なカテゴリ間の関係性を捉えることができると考えられる．英語と日本語のデータの両方を用いた カテゴリ A とカテゴリ B の距離 $dis(A,B)$ は

$$dis(A,B) = \frac{\lambda * disJP(A,B) + (1-\lambda) * disEN(A,B)}{2} \tag{5}$$

で計算される．式中の $disEN, disJP$ は英語のデータおよび日本語のデータだけで計算した カテゴリ間の距離を表し，λ は重み付けのための定数である．今回，$\lambda = 0.5$ とした．

5 評価実験

5.1 実験に使用したアンドロイドアプリのデータ

実験に使用したアプリのデータの内訳は，以下の表のとおりである．この内の日本語のデータ 10 件をランダムに抽出して，

言語	データ量
日本語	9416
英語	16519

表 1 データの構成

そのアプリケーションが保有する権限に対する機能の説明を人手で作成し，検証用データとした．今回対象とする権限はマイク，位置情報，カメラの 3 種類に絞った．それらの検証用データを用いて，システムによって抽出された文と人手で作られた

文との比較を行い提案手法を評価する．実験に使用した検証用データの例を図 1 に示す．

アプリケーション名	マップ - ナビ、乗換案内
カテゴリ	旅行＆地域
説明文	Google マップがあればお出かけは簡単。時間も節約できます。220 の国と地域を広くカバーする地図と数億のお店や場所に関する詳しい情報をご覧いただけます。リアルタイムの GPS ナビ…
権限1(位置情報)	○
権限2(マイク)	○
権限3(カメラ)	○
回答:位置情報(ベースライン)	道案内を行う
回答:位置情報(提案手法)	GPSで現在地を取得して道案内
回答:マイク(ベースライン)	…
回答:マイク(提案手法)	…
回答:カメラ(ベースライン)	…
回答:カメラ(提案手法)	…
正答:位置情報	GPSを用いたナビゲーション
正答:マイク	音声入力を用いた検索
正答:カメラ	カメラを用いた…

図 1　データの例

5.2　定量評価の方法

定量評価は Liu らの先行研究 [6] と同様の 2 種類の方法で行われる．

5.2.1　Jaccard 係数

Jaccard 係数を用いた定量評価は以下の手順で行われる．

（1） 提示された説明文を構成する単語と正解の文を構成する単語で Jaccard 係数を計算する．

（2） これを提示された各説明文に行い，平均をとる．

（3） 各アプリケーションについてこれを行い合計をとる．

5.2.2　Word2vec

Word2vec を用いた方法は以下の手順で行われる．

（1） 提示された説明文を構成する単語と正解の文を構成する単語を Word2vec でベクトル化する．

（2） 各ベクトル間でコサイン類似度をとり平均をとる．

（3） これを提示された各説明文に行い，平均をとる．

（4） 各アプリケーションについてこれを行い合計をとる．

5.3　ベースライン

ベースラインには，検索の段階に先行研究で用いられた CLAP [6] と同様の検索アルゴリズムを使用したものを用いる．これを提案手法を適用したデータに対して同様に使用し，抽出された文章を評価することでベースラインとする．

5.4　実 験 結 果

実験結果を表 2 に示す．

重み付けの λ は先行研究の方法に関しては 0.4, 0.4, 0.1, 0.1 とし, 提案手法の方法に関しては $\lambda_1 = \lambda_2$ =0.004 とした．

6　考　　察

Jacard 係数を用いた指標において先行研究は 0.0143 という数値を記録したのに対して翻訳拡張は 0.0074 という値を記録した．よって翻訳拡張はむしろマイナスに働いているということがわかる．これは翻訳によって生成される日本語の説明文と元から日本語である説明文において使われる語が異なるからであると考えられる．

一方，Word2vec を用いた指標において，先行研究の結果が 0.0104 となったが，翻訳拡張の結果の方は 0.0164 と先行研究よりも大きくなっている．このことから翻訳拡張は意味を正確に捉えるという点においては有効に働いていると考えられる．

また，両評価指標において先行研究のままの値は 0.0143,0.0104 なのに対して，カテゴリの類似度を用いたときは 0.0273,0.0289 となり，先行研究を上回った．これは，カテゴリの特徴を得る際に英語のデータを用いることは有効であるということを示すと考えられる．しかし，今回用いた評価対象のアプリケーションにはカテゴリの偏りがある．そのため更なる検証が必要である．

7　結　　論

本研究では,Anrdoid アプリケーションが権限を要求する理由への補足説明を試みた．そのために，既存のアプリケーションの説明文から対象のアプリケーションの説明に関係ありそうなものを抽出し，付与することを行った．先行研究との差異としては類似度の計算に意味情報を利用する，カテゴリ間の類似度の計算に複数の言語のデータを用いる，権限に対応する語を利用するといったことが挙げられる．結果として，先行研究よりも良い結果となったが，検証対象が少ないため更なる検証が求められる．

今後は以下などを行っていく．

- 候補文の生成方法を検討する
- 正解データの作成を数人で行うことでより正確な正解とする
- ユーザからアンケートをとる定性評価を行う

手法	先行研究	1	2	3	1,2	1,3	2,3	1,2,3
Jaccard 係数	0.0143	0.0074	0.0273	**0.0669**	0.0217	0.0297	0.0527	0.0268
Word2vec	0.0104	0.0164	0.0289	0.0683	0.0224	**0.0763**	0.0599	0.0657

表 2　実験結果 表の 1, 2, 3 はそれぞれ翻訳拡張の利用，カテゴリ類似度の拡張を利用，検索のステップにおけるアプリケーションの説明文の類似度の計算方法，および文の決定におけるランク付けの方法を先行研究と変更 に対応する

文　　献

[1] Google announces over 2 billion monthly active devices on android. https://www.theverge.com/2017/5/17/15654454/android-reaches-2-billion-monthly-active-users.

[2] Number of available applications in the google play store from december 2009 to june 2018. https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/.

[3] Record levels of app downloads & app store consumer spend in q4 2017. https://www.appannie.com/en/insights/market-data/app-downloads-consumer-spend-q4-2017/.

[4] Yisha Luo Weijia Jia Dong Xuan Nan Xu, Fan Zhang and Jin Teng. *Stealthy Video Capturer: A New Video-based Spyware in 3G Smartphones*. ACM, 2009.

[5] Yahoo!乗換案内　無料の時刻表、運行情報、乗り換え検索. `https://play.google.com/store/apps/details?id=jp.co.yahoo.android.apps.transit&hl=ja`.

[6] Xueqing Liu, Yue Leng, Wei Yang, ChengXiang Zhai, and Tao Xie. Mining android app descriptions for permission requirements recommendation. *2018 IEEE 26th International Requirements Engineering Conference (RE)*, pp. 147–158, 2018.

[7] S. E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pp. 232–241, New York, NY, USA, 1994. Springer-Verlag New York, Inc.

[8] Wei Yang William Enck Tao Xie Rahul Pandita, Xusheng Xiao. *WHYPER: Towards Automating Risk Assessment of Mobile Applications*. USENIX, 2013.

[9] Xinyi Zhang Yan Chen Tiantian Zhu Zhong Chen Zhengyang Qu, Vaibhav Rastogi. *AutoCog: Measuring the Description-to-permission Fidelity in Android Applications*. ACM, 2014.

[10] Alessandra Gorla Ilaria Tavecchia Florian Gross Andreas Zeller. Checking app behavior against app descriptions. *ICSE*, 2014.

[11] Mohammad Nauman, Sohail Khan, and Xinwen Zhang. Apex: Extending android permission model and enforcement with user-defined runtime constraints. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '10, pp. 328–332, New York, NY, USA, 2010. ACM.

[12] S. Y. Yerima, S. Sezer, G. McWilliams, and I. Muttik. A new android malware detection approach using bayesian classification. pp. 121–128, March 2013.

[13] Erika Chin, Adrienne Porter Felt, Vyas Sekar, and David Wagner. Measuring user confidence in smartphone security and privacy. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pp. 1:1–1:16, New York, NY, USA, 2012. ACM.

[14] Jialiu Lin, Shahriyar Amini, Jason I. Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pp. 501–510, New York, NY, USA, 2012. ACM.

[15] Timothy Vidas, Nicolas Christin, and Lorrie Faith Cranor. Curbing android permission creep. In *In W2SP*, 2011.

[16] Rada Mihalcea, Courtney Corley, and Carlo Strapparava. Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1*, AAAI'06, pp. 775–780. AAAI Press, 2006.

[17] Tom Kenter and Maarten de Rijke. Short text similarity with word embeddings. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pp. 1411–1420, New York, NY, USA, 2015. ACM.

[18] T. Sakai. The use of external text data in cross-language information retrieval based on machine translation. In *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 6, pp. 6 pp. vol.6–, Oct 2002.

[19] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004.