

IEEE Software

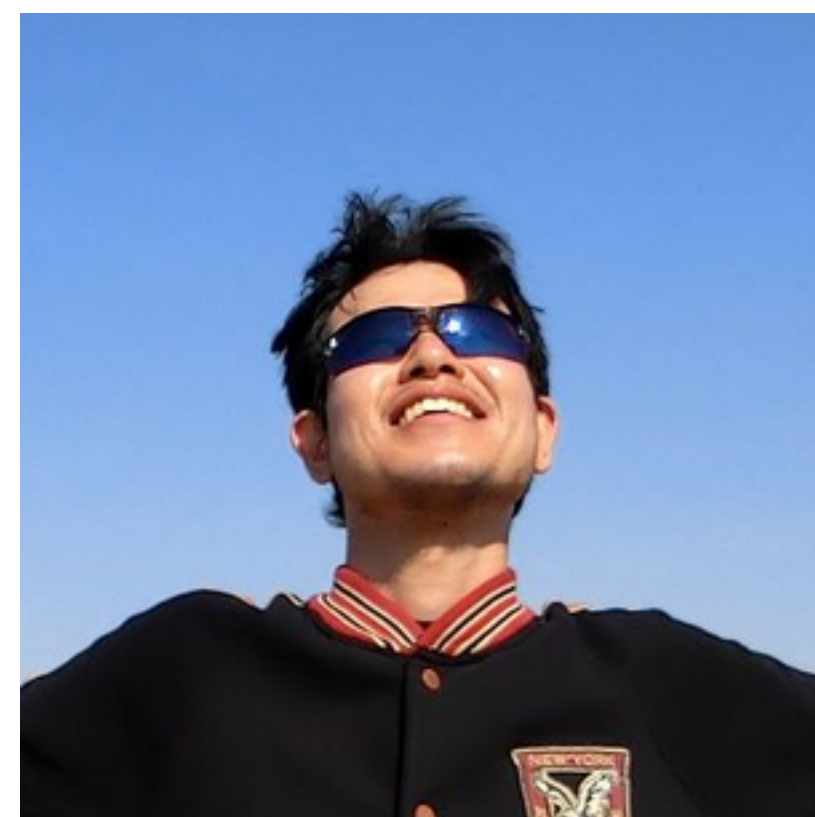
The 33rd IEEE International Conference on
Software Maintenance and Evolution (ICSME'17)

Bug or Not?

Bug Report Classification using N-Gram IDF



P. Terdchanakul



H. Hata



P. Phannachitta



K. Matsumoto

紹介者：柏 祐太郎

- **BugかRequestの分類は大変**
 - 最近ではTopic modelingによる分類が精度が良い
- **LDAなどのTopic Modelはパラメータ設定が大変**
 - トピック数やハイパーパラメータなど
- **本論文では、手間のかからない方法を提案**
 - N-gram IDF を利用する

IDF

- IDFは語の重み付けに有効

$$IDF(t) = \log \frac{|D|}{df(t)}$$

ex.

$$\log \left(\frac{\text{文書数}}{\text{単語}t\text{が出現する文書数}} \right)$$

- 単語のレア度
- 簡潔性やロバスト性が高い
- 一方, N-gramには使えないことが知られている
 - 単語間の繋がりの強さに反した重み付けをする

N-gram IDF

- IDF をコルモゴロフ複雑性に基づく情報距離と定義
 - 単語のレア度 (IDF) から単語間結合度 (MED) を引く
 - 単語間結合度 (MED) とは, 共起関係を利用した単語の繋がり of 不自然度

$$IDF_{N\text{-gram}}(g) = IDF(\theta(g)) - MED(g)$$

$$= \log \frac{|D|}{|\mu(g)|} - \log \frac{|\mu(g)|}{|\phi(g)|} = \log \frac{|D| \cdot df(g)}{df(\theta(g))^2}$$

- 単語と複合語に対する重みの比較が可能
 - 任意の長さの特徴語を言語処理技術なくして扱える

適用方法

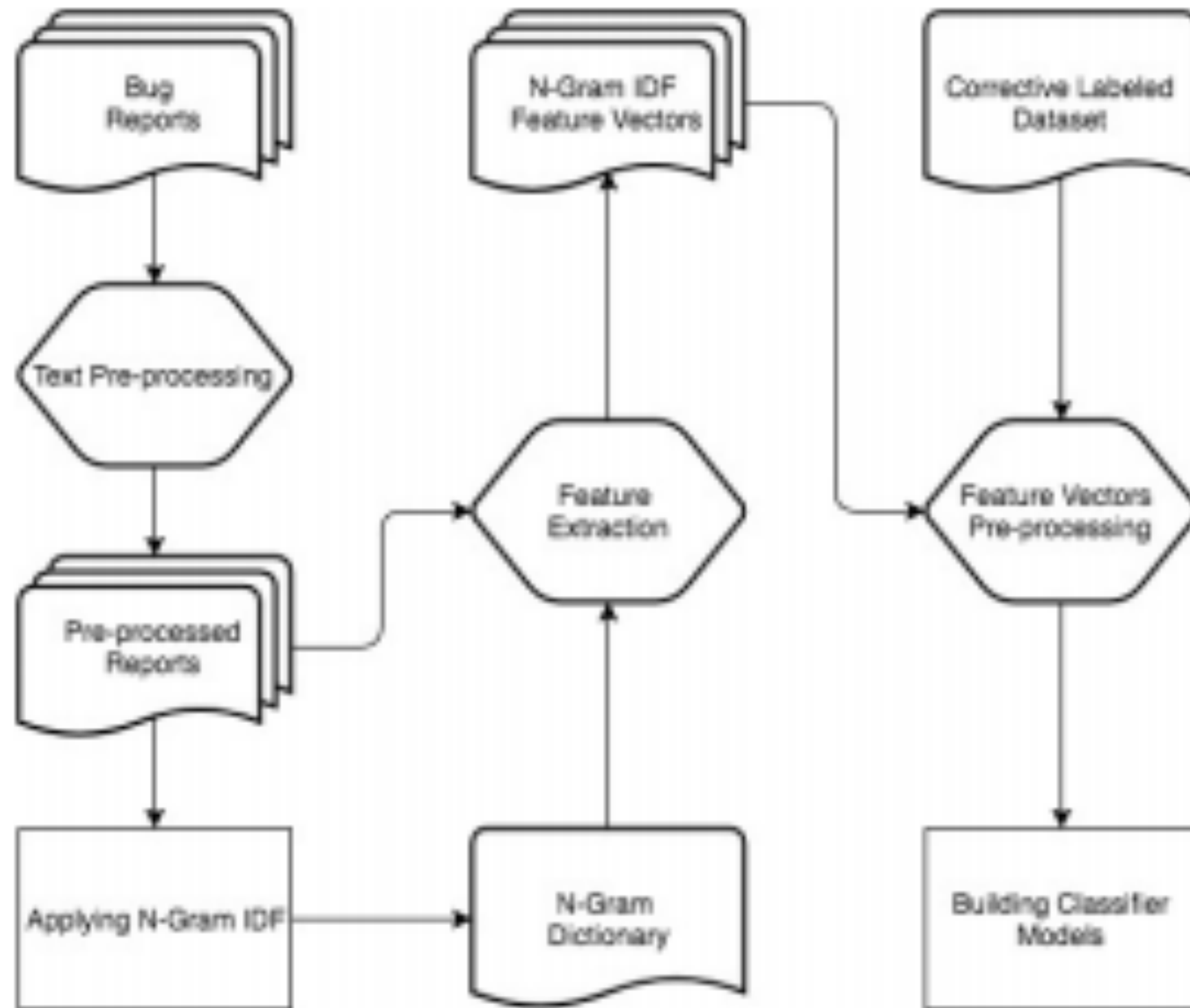


Fig. 1. Overview of our automatic bug reports classification model

1. 前処理

- (1) バグレポートをパース
- (2) ゴミ除去

2. N-gram 辞書作成

- (1) N-gram IDFの適用

3. 特徴ベクトルを抽出

- (1) 単語の出現回数をカウント
(一度しか出現しない単語は削除)

4. 特徴選択 (フィルタリング)

- (1) 相関の強い単語は削除(10-fold cross val.)
- (2) 分類に効く単語を選択(Chi-squared stats)

5. モデルの構築 (ロジスティック・RF)

実験

- バグレポートとリクエストを分類し,
 - F-measure : Recall/Precisionの調和平均で比較
 - 10-fold cross validation と 学習・テストを実施
- 比較対象
 - LDAを用いてテキスト処理した手法（既存手法） [1]
- データセット
 - 既存手法で使われていたものを流用
 - プロジェクト : HTTPClient, Jackrabbit, Lucent

[1] N. Pingclasai, H. Hata, and K. Matsumoto, “Classifying bug reports to bugs and other requests using topic modeling,” in Proceedings of 20th Asia-Pacific Software Engineering Conference (APSEC), vol. 2, pp. 13– 18, IEEE, 2013.

• 統計情報

- 単語数：58,000 ~ 530,000 (特徴抽出後: 40~200)

• 分類結果

TABLE II

F-MEASURE COMPARISON TABLE BETWEEN TOPIC-BASED AND N-GRAM IDF-BASED CLASSIFICATION MODELS BY 10-FOLD CROSS-VALIDATION

	Logistic Regression				Random Forest			
	HTTPClient	Jackrabbit	Lucene	Cross Project	HTTPClient	Jackrabbit	Lucene	Cross Project
Topic-based	0.739	0.744	0.766	0.724	0.721	0.717	0.756	0.712
N-gram IDF-based	0.805	0.805	0.884	0.814	0.814	0.771	0.823	0.792

TABLE III

F-MEASURE COMPARISON TABLE BETWEEN TOPIC-BASED AND N-GRAM IDF-BASED CLASSIFICATION MODELS BY TRAINING - TESTING SETUP

	Logistic Regression				Random Forest			
	HTTPClient	Jackrabbit	Lucene	Cross Project	HTTPClient	Jackrabbit	Lucene	Cross Project
Topic-based	0.516	0.511	0.562	0.592	0.494	0.514	0.566	0.542
N-gram IDF-based	0.687	0.646	0.731	0.658	0.673	0.628	0.685	0.674

• 出てきた単語

- gtf: 全てのドキュメントにおける単語の出現回数
- df: 単語が出現したドキュメント数

N-gram	Implication	gtf	df
nullpointerexception	Reports with a null pointer exception error is related to BUG.	252	140
how to reproduce	Reports with reproduce step should related to BUG.	17	14
improvement	Improvement discussion should not related to BUG.	464	295
demonstrate the problem	Discussion with demo of problem should related to BUG	11	9
performance test	Performance test discussion should not related to BUG	50	32

⚙️ N-gram IDFが新しい・流行りそう

⚙️ キーワード抽出に使えるそう
ハイ インパクトバグの分類など