

# 中間報告

2021/10/5

楠本研究室 M2

市川 直人



# 研究背景 | App review の分類

- アプリケーションレビューは開発者にとって重要な情報源である

## アプリレビューの種類

### - バグ報告

After the new update, my mobile freezes

### - 機能要求

It would be great if we could copy and paste text

### - 評価

Great app I love itttt

### - 無意味

:)

# 研究背景 | App review の分類

- アプリレビューを分類する研究が行われている
  - 自然言語処理
  - 機械学習（教師あり学習）

## Bug Report, Feature Request, or Simply Praise? On Automatically Classifying App Reviews

Walid Maalej  
University of Hamburg  
Hamburg, Germany  
maalej@informatik.uni-hamburg.de

Hadeer Nabil  
University of Hamburg  
Hamburg, Germany  
nabil.hadeer@gmail.com

## Classifying Multilingual User Feedback using Traditional Machine Learning and Deep Learning

Christoph Stanik, Marlo Haering and Walid Maalej  
University of Hamburg  
Hamburg, Germany  
{stanik, haering, maalej}@informatik.uni-hamburg.de

## Automatic Classification of Apps Reviews for Requirement Engineering: Exploring The Customers Need from Healthcare Applications

1<sup>st</sup> Nadeem Al Kilani  
MSc Software Engineering  
Birzeit University  
nadeemkilani@gmail.com

2<sup>nd</sup> Rami Tailakh  
MSc Applied Computing  
University of Bedfordshire  
rami.azmi@gmail.com

3<sup>rd</sup> Abualsoud Hanani  
Assistant Professor  
Birzeit University  
ahanani@birzeit.edu

**Abstract**—App stores like Google Play and Apple AppStore have over 3 Million apps covering nearly every kind of software and service. Billions of users regularly download, use, and review these apps. Recent studies have shown that reviews written by the users represent a rich source of information for the app vendors and the developers, as they include information about bugs, ideas for new features, or documentation of released features. This paper introduces several probabilistic techniques to classify app reviews into four types: bug reports, feature requests, user experiences, and ratings. For this we use review metadata such as the star rating and the tense, as well as, text classification, natural language processing, and sentiment analysis techniques. We conducted a series of experiments to compare the accuracy of the techniques and compared them with simple string matching. We found that metadata alone results in a poor classification accuracy. When combined with natural language processing,

U  
qua  
con  
Wit  
[16]  
filt  
A  
rev  
to t  
and  
fou  
app  
beh

and  
got  
cts.  
per  
ble  
em  
nal-  
any  
ses  
ies.  
line  
lish  
dur  
can  
we  
cial

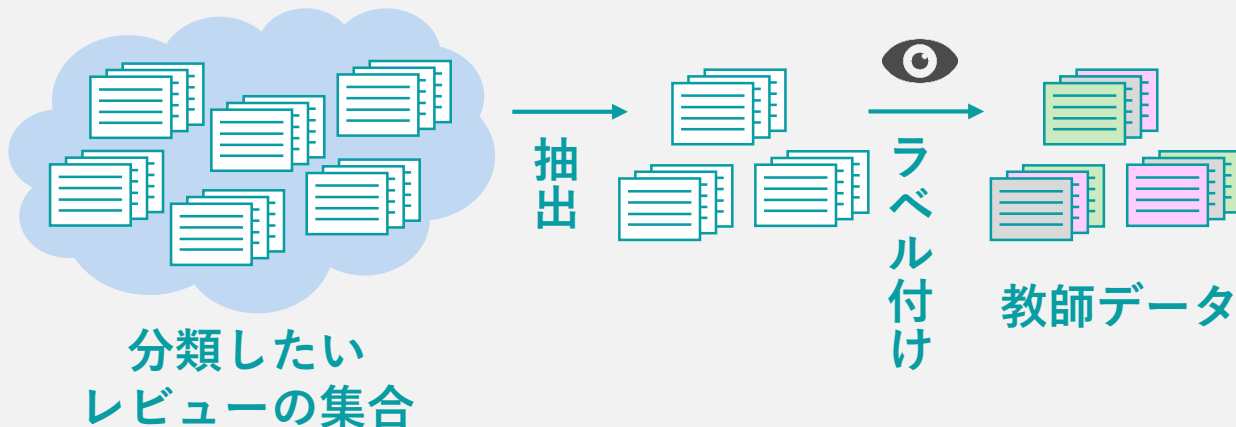
software product or service (e.g., “Since the last update the app crashes upon start”). We define inquires as user feedback that asks for either new functionality, an improvement, or requests information for support (e.g., “It would be great if I could invite multiple friends at once”). We consider user feedback as irrelevant if it does not belong to problem reports or inquires (e.g., “I love this app”).

To fulfill our objective, we employ supervised machine learning fed with crowd-sourced annotations of 10,000 English and 15,000 Italian tweets from telecommunication Twitter support accounts, and 6,000 annotations of English app reviews. We apply best practices for both machine learning approaches (traditional and deep learning) and report on a benchmark.

**Preliminary results.** Our preliminary results show that, within

# 研究背景 | 既存研究の分類手法

## データの用意



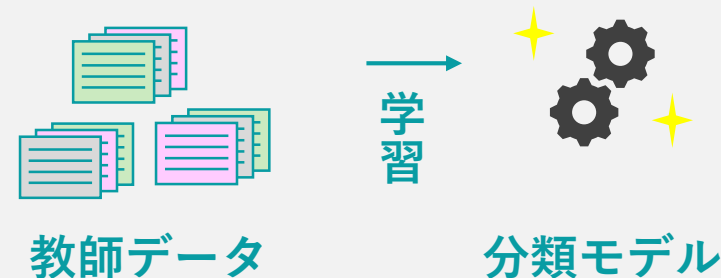
## 本研究での焦点

実際に**開発者**が  
レビューを分類  
する際の大きな  
ハードル

## 自然言語処理



## 分類モデル構築

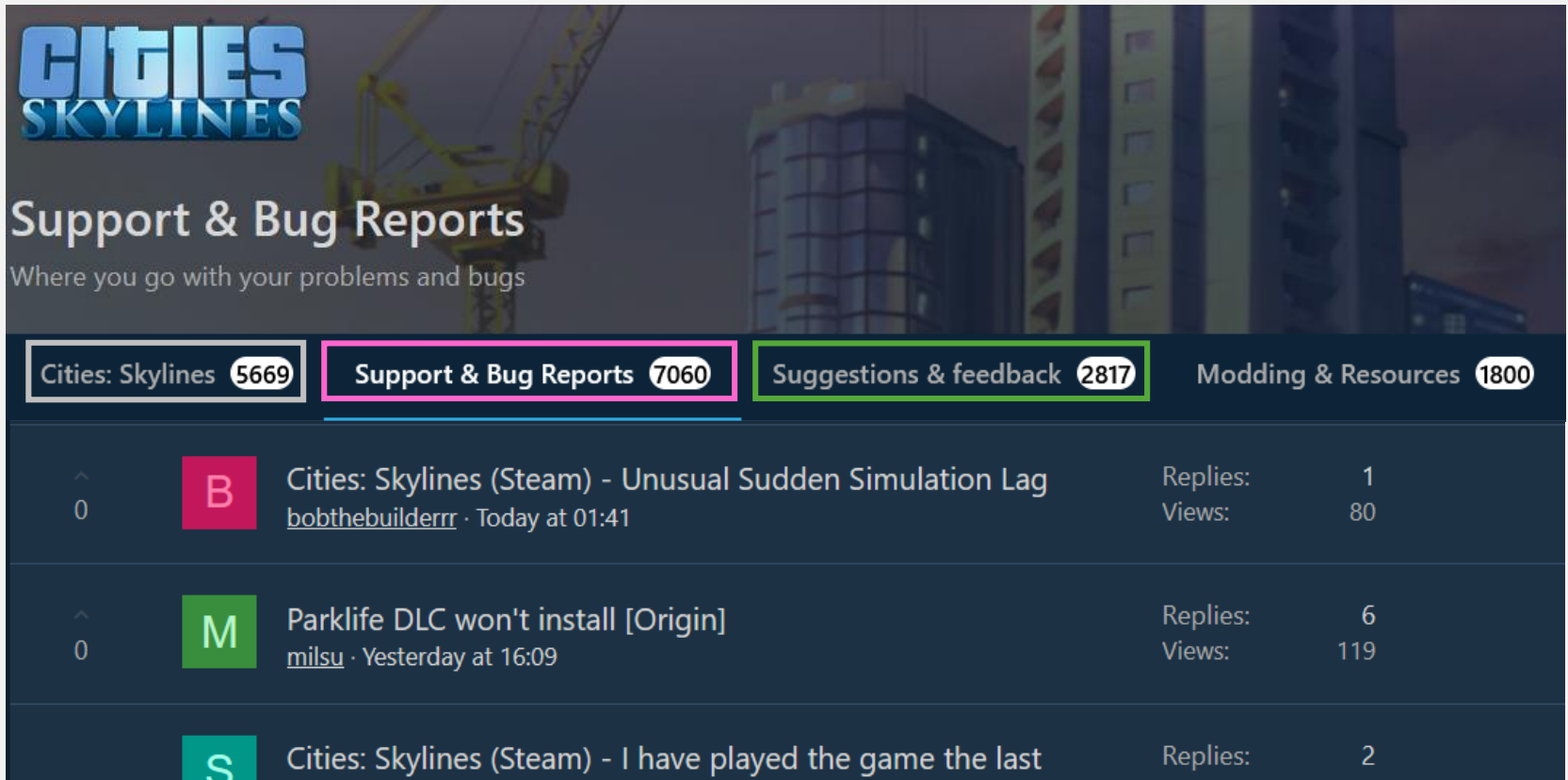


## 既存研究での焦点

# フォーラムについて

## フォーラム

- エンドユーザの議論の場
- 各トピックがカテゴリーごとに分かれている
- 主にゲームで盛んな文化



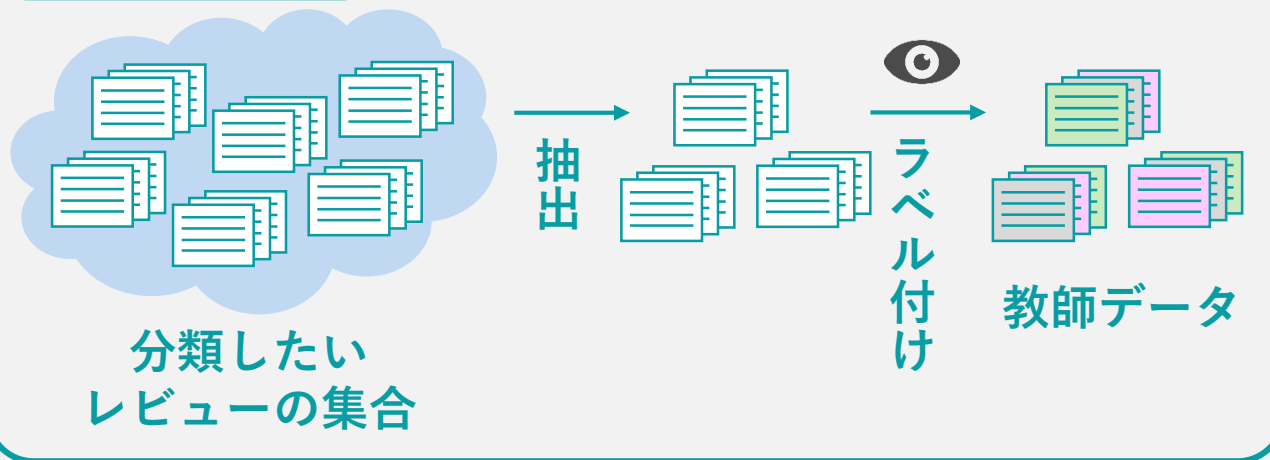
The screenshot shows the 'Cities: Skylines' forum interface. At the top, the 'CITIES SKYLINES' logo is visible. Below it, the section is titled 'Support & Bug Reports' with the subtitle 'Where you go with your problems and bugs'. A navigation bar contains four categories: 'Cities: Skylines' (5669), 'Support & Bug Reports' (7060), 'Suggestions & feedback' (2817), and 'Modding & Resources' (1800). The 'Support & Bug Reports' category is highlighted with a pink border. Below the navigation bar, a list of forum posts is displayed. Each post includes an icon, a title, the user name, the time, and the number of replies and views.

Icon	Title	User	Time	Replies	Views
B	Cities: Skylines (Steam) - Unusual Sudden Simulation Lag	bobthebuilderrr	Today at 01:41	1	80
M	Parklife DLC won't install [Origin]	milsu	Yesterday at 16:09	6	119
S	Cities: Skylines (Steam) - I have played the game the last			2	

# 研究背景 | 既存研究の分類手法

再掲

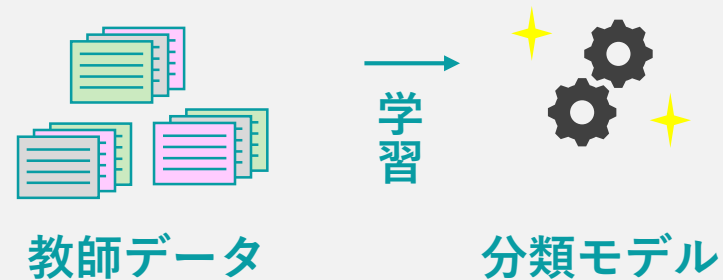
## データの用意



## 自然言語処理



## 分類モデル構築

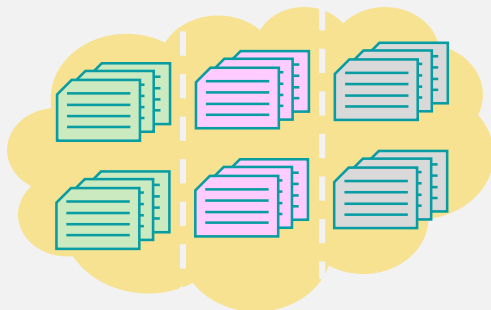


# 提案手法

## データの用意



分類したい  
レビューの集合

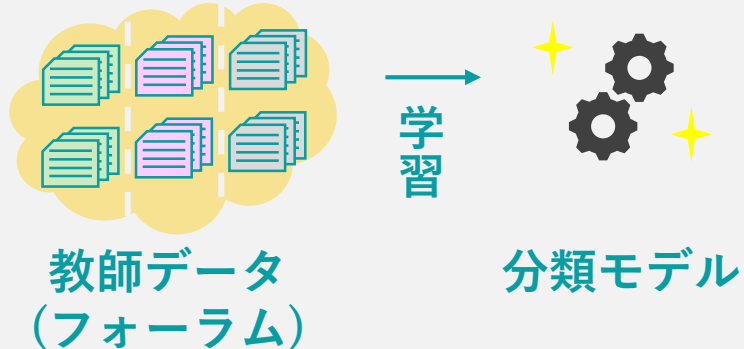


フォーラム

## 自然言語処理



## 分類モデル構築



# 研究概要

## フォーラムを教師データとした レビューの分類手法の提案

### 貢献

**アプリ開発者**がレビュー分類を行う際、  
目視調査を行わずに分類ができる

- Steam上のフォーラムを持つゲームへの  
レビューを対象に調査
- レビューを以下の3種類に分類
  - バグ報告
  - 機能要求
  - その他
- 分類精度は低下すると予想



Support & Bug Reports 7060

Suggestions & feedback 2817

Cities: Skylines 5669



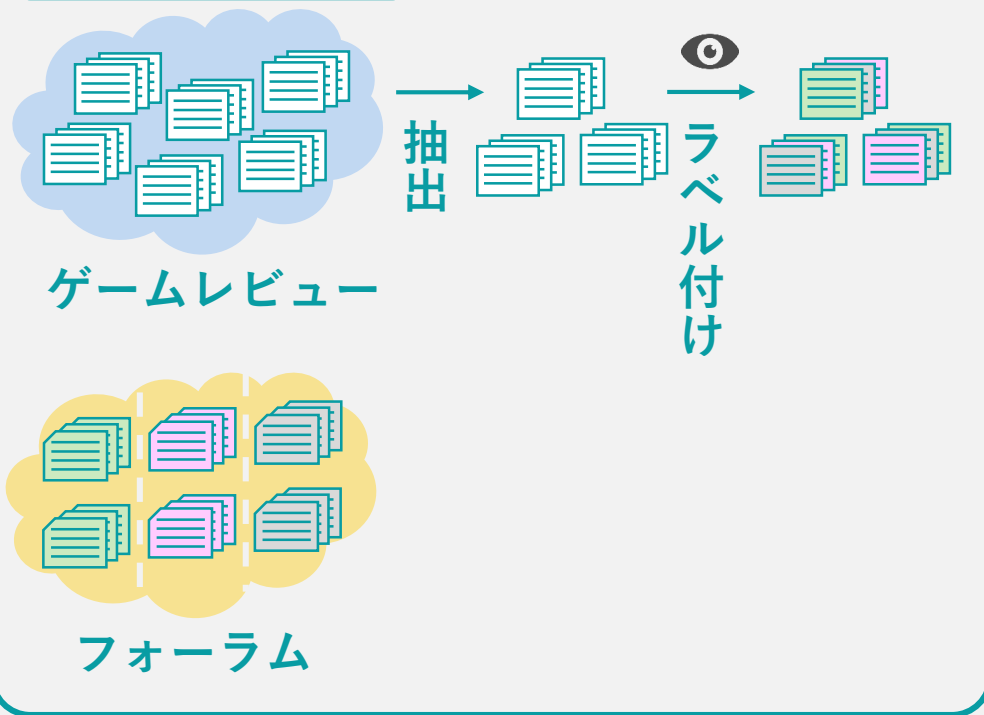
# Research Question

RQ1. 同一のタイトルのフォーラムを教師データとした場合の分類精度はどの程度か？

RQ2. 異なるタイトルのフォーラムを教師データとした場合の分類精度はどの程度か？

# study design

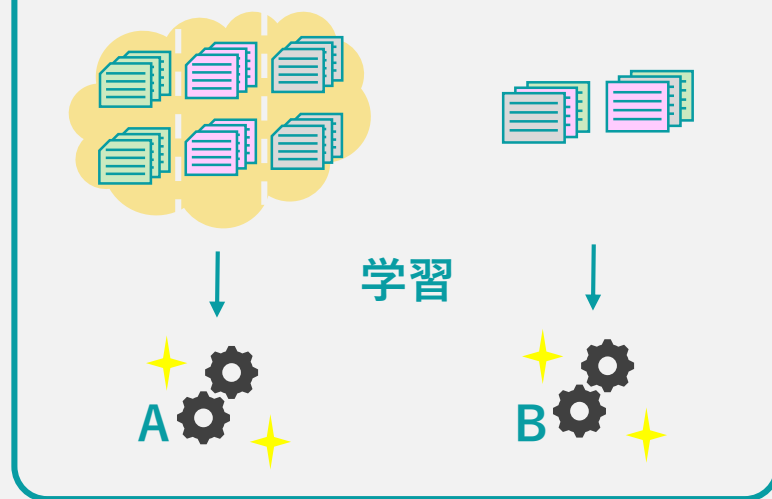
## 1. データの用意



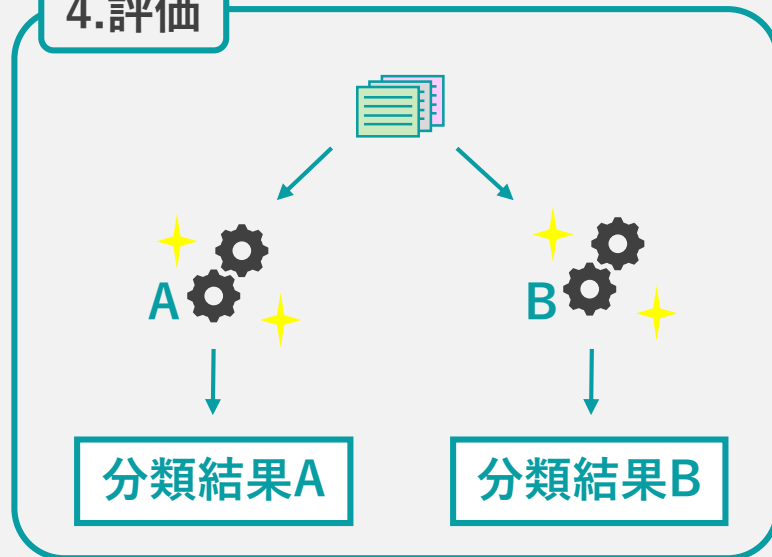
## 2. 自然言語処理



## 3. 分類モデル構築



## 4. 評価



# データの用意

- 対象となるタイトルの選定
  - Steamのおすすめ上位50件について、フォーラムの有無を調査したところ10件が該当
  - 上記10件からnタイトルを研究対象とする
- Steamレビューの取得
  - 対象タイトルの英語レビューを取得
  - Steamが公開しているAPIを利用
- フォーラムの取得
  - 対象タイトルのフォーラムから英語のトピックを取得
  - Webスクレイピングによって取得

# 自然言語処理 | 前処理

## 小文字化

- 文章を小文字に統一

## トークナイズ

- 文章を単語ごとに分割

## ストップワード除去

- 頻出する一般的な英単語を除去

## レマム化

- 各単語を見出し語に変換

i bought this game yesterday  
and it is full of bugs!

i/bought/this/game/yesterday  
/and/it/is/full/of/bugs/!

i bought this game yesterday  
and it is full of bugs!

i **buy** this game yesterday  
and it is full of **bug**!

buy game yesterday full of bug!

# 自然言語処理 | 分散表現の取得

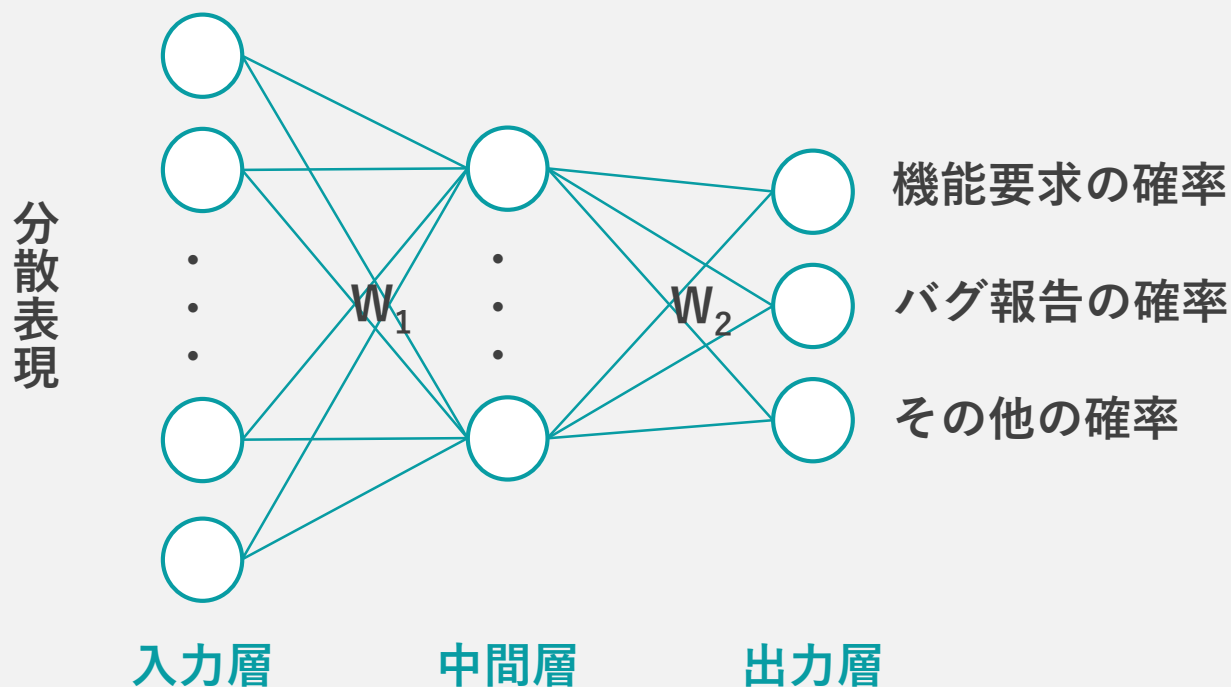
## BERT

- 2018年Googleが発表
- 穴あきの文章から単語を予想するタスクを行う  
ニューラルネットワークモデルの重みを文章の分散表現として用いる
- 事前に巨大なコーパスで学習した結果を初期値として追加で学習する
- 次元数が語彙数に抛らず一定
- 文脈を考慮した分散表現

# 分類モデル構築

## ニューラルネットワーク

- ヒトの脳（ニューロン）を模した数理モデル



# 実験

## Cities: Skylines のレビュー1175件をラベル付け

レビュー		
バグ報告	機能要求	その他
108	67	1000



7 : 3



訓練用レビュー		
バグ報告	機能要求	その他
80	47	695

検証用レビュー		
バグ報告	機能要求	その他
28	20	305

## Cities: Skylines のフォーラムから約1万件のトピックを取得

フォーラム		
バグ報告	機能要求	その他
5166	1863	4468

# 実験結果

既存手法 review-review		予測した分類			適合率 precision	再現率 recall	F値 F1-score
		バグ報告	機能要求	その他			
真の分類	バグ報告	24	2	2	0.86	0.86	0.86
	機能要求	1	6	13	0.38	0.30	0.33
	その他	3	8	294	0.95	0.96	0.96

正解率 : 0.92

提案手法 forum-review		予測した分類			適合率 precision	再現率 recall	F値 F1-score
		バグ報告	機能要求	その他			
真の分類	バグ報告	22	2	4	0.48	0.79	0.59
	機能要求	1	13	6	0.32	0.65	0.43
	その他	23	26	256	0.96	0.84	0.90

正解率 : 0.82



# 今後の予定

日程	内容
10月	データ収集・実験
11月中旬	論文執筆開始
12月	SIGSS論文投稿
1月	SIGSS発表