

Применение методов машинного обучения для автоматической рекомендации рефакторинга “перемещение метода”

Веселов Иван Дмитриевич

Руководитель: Брыксин Тимофей Александрович

НИУ ВШЭ СПб

7 июня 2019

Рефакторинг “перемещение метода”

```
class A {  
    void foo() {  
    }  
  
    void method(B b) {  
        b.field = 0;  
        foo();  
    }  
}  
  
class B {  
    int field;  
}
```

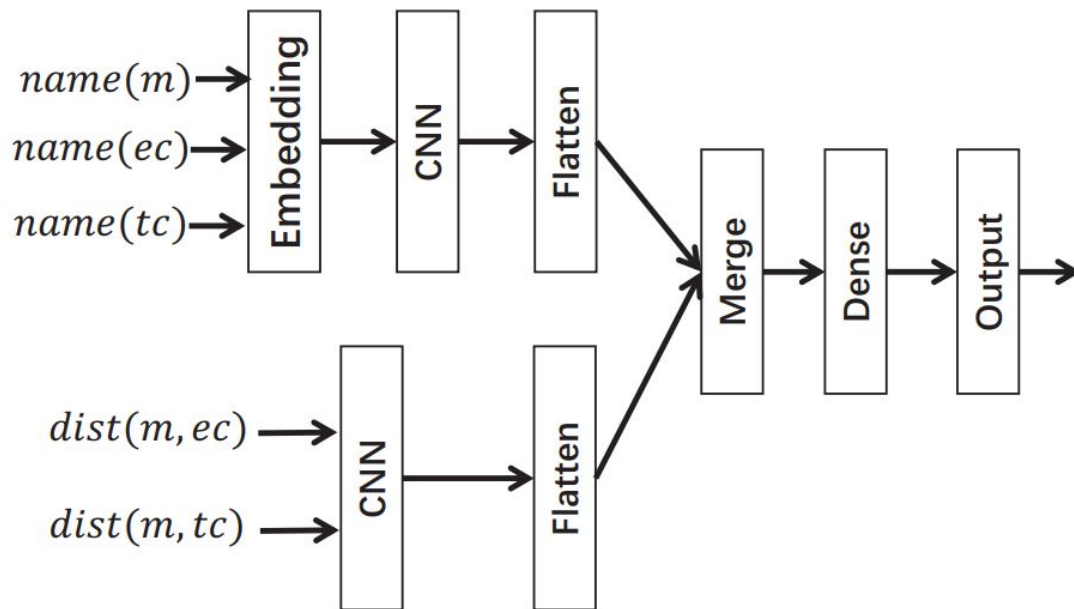


```
class A {  
    void foo() {  
    }  
}  
  
class B {  
    int field;  
  
    void method(A a) {  
        field = 0;  
        a.foo();  
    }  
}
```

Алгоритмы автоматической рекомендации

- Вычисление расстояния
 - Множества зависимостей
- Анализ текста
 - Идентификаторы и комментарии
 - LSA
- Анализ истории
 - Коммиты в git репозиториях
- Улучшение метрик качества
 - Coupling, cohesion...
- Машинное обучение
 - Word2vec
 - CNN

Машинное обучение



Цель и задачи

Разработка лучшего алгоритма, который использует **машинное обучение** для рекомендации рефакторинга “перемещения метода” для языка программирования **Java**.

Задачи:

1. Собрать **данные** для обучения и проверки модели
2. Придумать **модель**, предсказывающую необходимость рефакторинга
3. **Обучить** модель
4. **Сравнить** полученный подход с существующими решениями

Задача 1: существующие датасеты

- Recommending Move Method refactorings using dependency sets, **2013**
 - *V. Sales, R. Terra, L. F. Miranda and M. T. Valente*
 - 14 проектов, 395 методов
- On the diffuseness and the impact on maintainability of code smells, **2018**
 - *Fabio Palomba, et al*
 - Меньше ста уникальных методов
- Landfill: An Open Dataset of Code Smells with Public Evaluation, **2015**
 - *Fabio Palomba, et al*
 - 86 методов, платформа недоступна

Задача 1: генерация

- Поиск и перемещение всех возможных методов
- Утилита основана на IntelliJ Platform SDK
- Всего проектов: **1000** (в 485 ничего не найдено)
- Количество методов: **18 086**
- Количество точек в датасете: **38 320** (0.47 pos / 0.53 neg)

Sales V., Terra R., Miranda L.F., Valente M.T. 2013.
Recommending Move Method refactorings using
dependency sets
20th Working Conference on Reverse Engineering (WCRE)

Hui Liu, Zhifeng Xu, and Yanzhen Zou. 2018.
Deep learning based feature envy detection
In Proceedings of the 33rd ACM/IEEE International
Conference on Automated Software Engineering
(ASE 2018)

Задача 2: модель, векторизация

- Стандартные метрики работают плохо

- Количество методов в классе, глубина дерева наследования...
- Рекомендуется комбинировать с другими методами

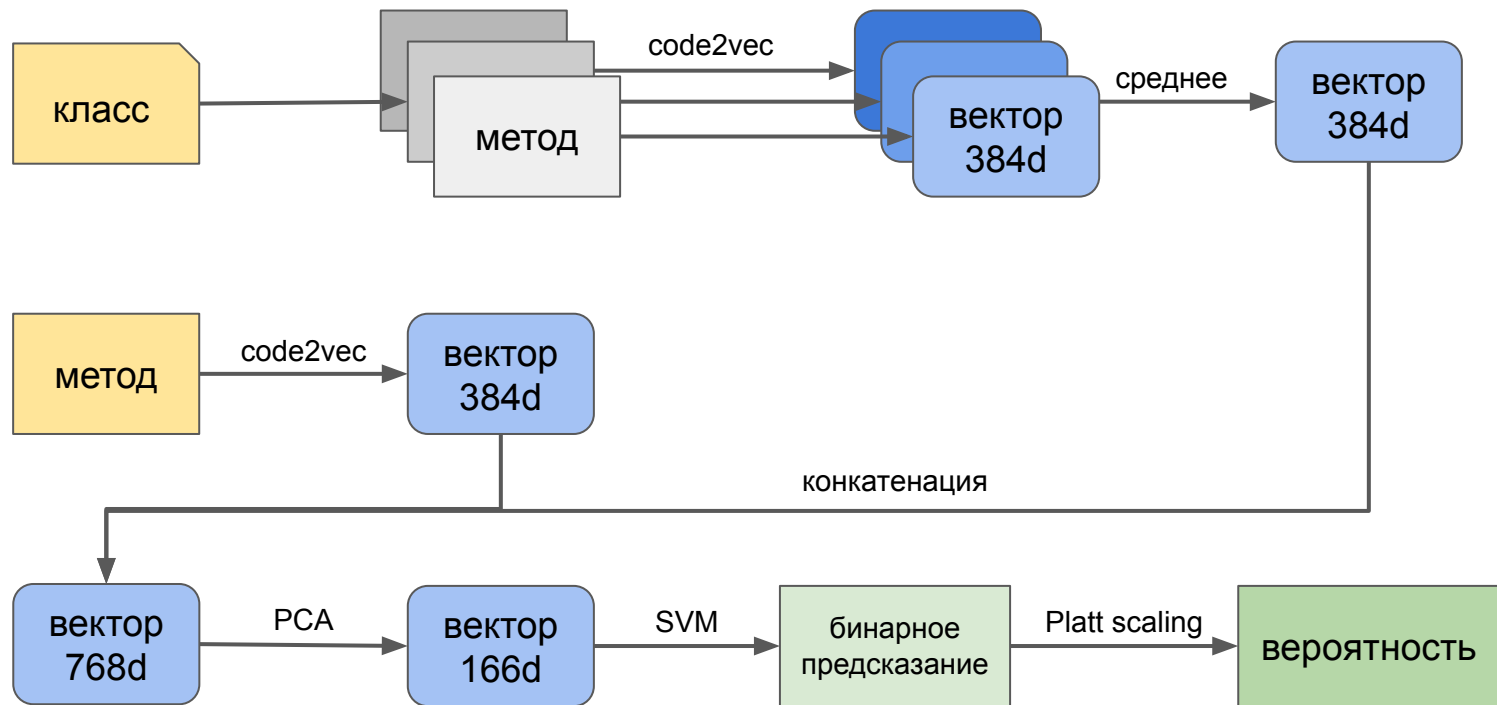
Di Nucci D., Palomba F., Tamburri D.A., Serebrenik A., De Lucia A. 2018 Detecting code smells using machine learning techniques: Are we there yet? 25th IEEE International Conference on Software Analysis, Evolution and Reengineering

- Обучение представлениям

- word2vec не предназначен для анализа кода
- Недавно была представлена code2vec
- Есть предобученная авторами модель

Alon U., Zilberstein M., Levy O., Yahav E. 2019 code2vec: Learning Distributed Representations of Code. Proc. ACM Program. Lang. 3, POPL

Задача 2: модель



Задача 3: результаты модели

	Не применять	Применять
Precision	0.798	0.788
Recall	0.785	0.801
F1-score	0.791	0.795
ROC AUC	0.793	

Разбиение собранного датасета

Обучение: 24 434 (30 проектов)

Валидация: 7 910 (40 проектов)

Тестирование: 8 124 (403 проекта)

Задача 4: сравнение

- JDeodorant

- Во многих статьях делают сравнение именно с этой работой
- Подходящая для сравнения инфраструктура

Fokaefs M., Tsantalis N., Chatzigeorgiou A. 2007. JDeodorant: Identification and Removal of Feature Envy Bad Smells. IEEE International Conference on Software Maintenance

- Qualitas.class: 111 проектов для Eclipse

- Требовались Eclipse проекты, которые компилируются

Terra R., Miranda L., Valente M., Bigonha R. 2013 Qualitas.class Corpus: A compiled version of the Qualitas Corpus. SIGSOFT Softw. Eng. Notes 38, 5

- Небольшой датасет

- Использовалась созданная утилита для переноса методов
- 49 методов, 2 проекта

Задача 4: сравнение

	Не применять		Применять	
	Данная работа	JDeodorant	Данная работа	JDeodorant
Precision	0.739	0.452	0.712	0.360
Recall	0.694	0.673	0.755	0.184
F1-score	0.716	0.541	0.733	0.243

Тест МакНемара: **p-value = 0.00106**

Результаты

- Создана утилита для генерации датасета
 - Поиск методов, которые можно переносить в другие классы
 - Перенос найденных методов
 - Представлена на IWOR 2019 как самостоятельный результат
 - Собран датасет из **18 086** методов
- Придумана и обучена модель на основе code2vec
 - code2vec хорошо справляется с рекомендацией рефакторинга
 - Значение ROC AUC равно **79%**
- Проведено сравнение с существующим решением
 - Сравнение проведено на небольшой выборке методов
 - Модель справляется лучше JDeodorant, F1-score выше на **17%**