

Система оценки качества инструментов автодополнения кода в IntelliJ IDEA

Тучина Анастасия Игоревна

группа 16.Б09-мм

научный руководитель: к.т.н., доцент Т.А. Брыксин

рецензент: программист ООО "Интеллиджей Лабс" В.И. Бибаев

СПбГУ

6 июня 2020 г.

Актуальность

- ▶ В настоящее время эффективность инструментов автодополнения кода оценивают в основном на синтетических данных
- ▶ Обычно данные генерируются случайно, не учитывая особенности использования инструментов людьми
- ▶ Требуется инструмент для сравнения алгоритмов автодополнения, позволяющий получить реалистичную оценку их качества

Мотивация

| Тип | Процент |
|------------------|---------|
| Пунктуация | 57,1% |
| Идентификатор | 32.1% |
| Ключевое слово | 10.8% |
| Числовой литерал | 0.5% |

Таблица: Доли токенов каждого типа в синтетическом датасете

▶ Разница в точности предсказаний моделей на синтетических и реальных данных может превышать 20%

V.J. Hellendoorn et al., When code completion fails: a case study on real-world completions, ICSE '19 Proceedings of the 41st International Conference on Software Engineering

Постановка задачи

Цель:

- ▶ Разработать систему для интеграции инструментов автодополнения кода с IntelliJ Platform и оценки их качества

Задачи:

- ▶ Проанализировать существующие алгоритмы автодополнения кода и способы оценки их качества
- ▶ Разработать инструментарий, позволяющий интегрировать инструменты автодополнения кода с IDE на основе IntelliJ Platform
- ▶ Собрать датасет на основе данных об использовании людьми инструментов автодополнения
- ▶ Разработать сервис, позволяющий оценивать качество инструментов
- ▶ Провести апробацию разработанных инструментов на нескольких существующих моделях автодополнения кода

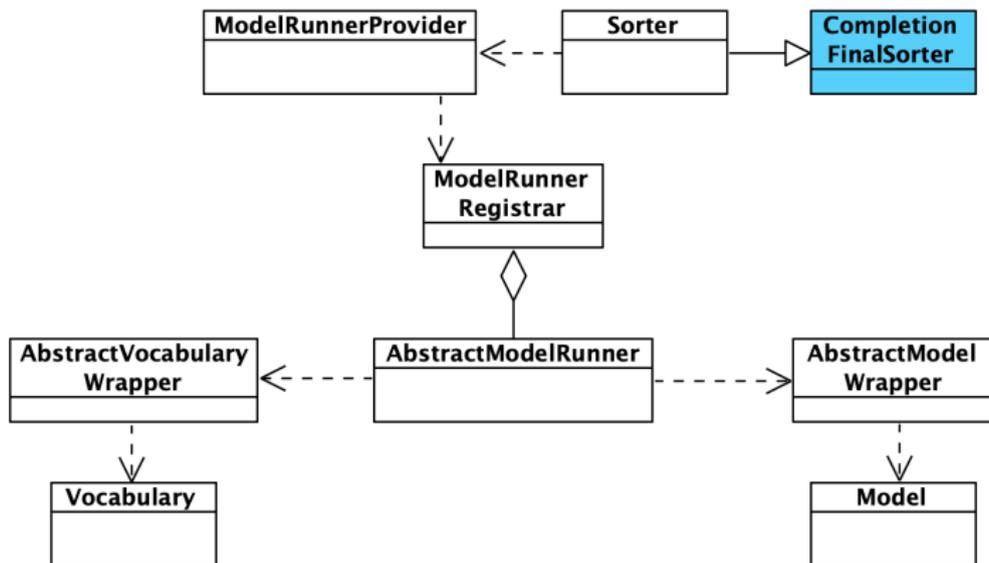
Современные методы автодополнения кода

- ▶ Методы на основе NLP
 - ▶ N-граммы
 - ▶ Синтез текста с помощью рекуррентных нейросетей и Transformer (TabNine)
- ▶ Методы, специфичные для кода
 - ▶ Structural feature selection
 - ▶ Дополнение только токенов определенного типа (API-recommenders)
 - ▶ Расширение NLP-моделей информацией из кода

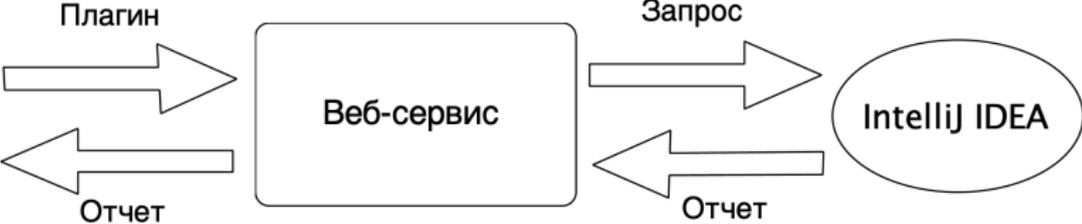
Методы оценки качества автодополнения

- ▶ Оценка на синтетических запросах
 - ▶ Удаление случайного токена из корпуса
 - ▶ Удаление токенов определенных типов из корпуса
 - ▶ Иногда токен удаляется только частично
- ▶ Оценка на пользовательской истории

Инструментарий для интеграции моделей



Веб-сервис



Датасет

- ▶ 4 крупных популярных репозитория с GitHub на Java
- ▶ Генерация сессий с помощью IntelliJ Completion Benchmark Plugin

| Имя репозитория | Описание | Количество строк кода на Java | Количество звезд |
|--------------------|---------------------------------------|----------------------------------|---------------------|
| intellij-community | IDE для Java и Kotlin | 179 620 948 | 9 877 |
| jenkins | CI инструмент | 10 158 027 | 15 547 |
| languagetool | Инструмент для проверки грамматики | 6 586 085 | 3 685 |
| deeplearning4j | Библиотека для машинного обучения | 34 706 214 | 11 645 |

Таблица: Выбранные репозитории

Генерация данных

Сессия:

- ▶ Введенный префикс
- ▶ Список рекомендаций (выдача)
- ▶ Ожидаемый результат
- ▶ Уникальный идентификатор

Метаинформация:

- ▶ Длина префикса
- ▶ Позиция корректного предложения (или -1)
- ▶ Тип токена
- ▶ Уникальный идентификатор

Параметры датасета

Выбор данных:

- ▶ Доля сессий из каждого репозитория
- ▶ Доля использованных сессий с определенной длиной префикса

Метрики:

- ▶ Top 1
- ▶ Top 5
- ▶ Средняя позиция

Апробация

| Префикс | | | | |
|-----------------|---------------------|-------|-------|-------|
| | | 0 и 1 | 2 | 3 |
| Метрика | | | | |
| Топ 1 | Ожидаемое значение | 0.609 | 0.762 | 0.782 |
| | Полученное значение | 0.624 | 0.769 | 0.799 |
| Топ 5 | Ожидаемое значение | 0.762 | 0.91 | 0.91 |
| | Полученное значение | 0.77 | 0.907 | 0.921 |
| Средняя позиция | Ожидаемое значение | 3.694 | 1.695 | 1.399 |
| | Полученное значение | 3.973 | 1.8 | 1.431 |

Таблица: Значения метрик на собранном датасете для стандартного автодополнения в IntelliJ IDEA

Результаты

- ▶ Проведен анализ существующих алгоритмов автодополнения кода и способов оценки их качества
- ▶ Разработан инструментарий для интеграции инструментов автодополнения кода с IntelliJ Platform
- ▶ Разработан прототип сервиса, позволяющий оценивать качество инструментов
- ▶ Собран датасет на основе значений нескольких полученных из пользовательской статистики IntelliJ IDEA метрик
- ▶ Собранный датасет был протестирован на других типах автодополнения IntelliJ IDEA