

УДК 004.852

ПРОГНОЗИРОВАНИЕ СТАТУСА ЗАВЕРШЕНИЯ BPMN-ПРОЦЕССОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Найденкова Е.Д. (Университет ИТМО)

Научный руководитель – к.т.н., доцент Дергачев А.М.
(Университет ИТМО)

Аннотация В данной работе рассматривается метод, благодаря которому можно получить прогноз об успешности или возможной ошибке в протекающем BPMN-процессе с помощью методов машинного обучения. Решение предлагается для приложений, основанных на BPMN-движке Camunda.

Введение.

Сегодня всё большую популярность приобретают методы машинного обучения. Они используются в различных сферах для решения задач классификации, кластеризации, регрессии и многих других. Применение же методов машинного обучения в приложениях, основанных на BPMN-процессах, всё ещё недостаточно изучено, хотя это позволило бы добиться ещё большей автоматизации как в процессе моделирования бизнес-процессов, так и во время выполнения самого процесса. В информационных системах, которые используются огромным количеством пользователей и состоят из сложных и долгих бизнес-процессов, важно как можно раньше узнать о неполадках в процессах, чтобы не потерять клиентов. Прогнозирование статуса завершения BPMN-процесса с помощью машинного обучения позволило бы решить эту проблему.

Основная часть.

Так как целью разрабатываемой системы является определение того, что BPMN-процесс завершится успешно или нет, то первым шагом нужно определить признаки, по которым можно сделать такой прогноз.

Был выявлен набор параметров, характеризующих текущее состояние BPMN-процесса и определяющих его работоспособность:

- Зафиксированные BPMN-движком инциденты;
- Время выполнения процессов;
- Время выполнения активити;
- Переменные, участвующие в вызовах методов из блока expression;
- Переменные, сохранённые в процессе.

Дальше была описана обобщённая модель BPMN-процесса, которая состоит из трёх сущностей: процесс, активити и инцидент. Характеристики этих сущностей схожи, для процессов это:

- processInstanceId – уникальный идентификатор экземпляра процесса;
- processId – название процесса;
- state – состояние процесса (завершён, выполняется, остановлен, прерван ошибкой);
- businessKey – бизнес ключ экземпляра процесса;
- startTime – время старта процесса;
- endTime – время окончания процесса;
- variables – переменные, которые можно сохранять в экземпляре процесса.

Система состоит из двух главных компонентов. Для их взаимодействия используется брокер сообщений. Первый — это модуль, который встраивается в приложения с BPMN-движком. Его задачей является преобразование данных, хранимых во внутренних базах данных BPMN-приложений, к формату модели определённой ранее и отправка в реальном времени этих событий во второе разрабатываемое приложение для их последующего анализа.

Второй компонент — это приложение, выполняющие прогноз о статусе завершения процесса. В качестве библиотеки для анализа данных и машинного обучения используется Weka. На этапе предварительной обработки данных и их изучения мы сохраняем значения, полученные из первого модуля. Важно, чтобы среди этих значений были успешно завершённые процессы, которые будут считаться эталонными. Для выполнения классификации использовался алгоритм Random forest. Так как мы хотим предсказать тип nominal, то библиотека выдаст точность классификации и матрицу ошибок. В качестве метода обучения был выбран статистический метод оценки модели - кросс валидация.

Выводы.

Результаты данной работы могут быть использованы в IT-компаниях, которые разрабатывают приложения на BPMN-движках. Прогноз о статусе завершения BPMN-процессов поможет сохранить вычислительные ресурсы за счёт удаления процессов, которые не будут завершены корректно. Также результаты работы демонстрируют, что методы машинного обучения можно успешно использовать с BPMN-процессами. Другой перспективной областью для их использования является поддержка принятия решений, которая позволит сократить количество операций, требующих участия человека.

Найденкова Е.Д. (автор)

Подпись

Дергачёв А.М. (научный руководитель)

Подпись