

65 ЛЕТ ГАЗОВАЯ ПРОМЫШЛЕННОСТЬ

№ 5

816 | 2021

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ И ПРОИЗВОДСТВЕННЫЙ ЖУРНАЛ, ИЗДАЕТСЯ С 1956 г.,
ВХОДИТ В ПЕРЕЧЕНЬ РЕЦЕНЗИРУЕМЫХ НАУЧНЫХ ИЗДАНИЙ ВАК



60 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ
ПЛАНИРОВАНИЯ ГЕОЛОГО-ТЕХНИЧЕСКИХ
МЕРОПРИЯТИЙ НА ГАЗОВЫХ
СКВАЖИНАХ

74 СПГ-СТРАТЕГИЯ
КАТАРА В КОНТЕКСТЕ
ДЕКАРБОНИЗАЦИИ МИРОВОЙ
ЭНЕРГЕТИКИ

92 ТЕХНОЛОГИЯ СЖИЖЕНИЯ ГАЗА
С ИСПОЛЬЗОВАНИЕМ В КАЧЕСТВЕ
ХЛАДАГЕНТА ГЕЛИЙСОДЕРЖАЩИХ
СМЕСЕЙ

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ МЕТОДА МАШИННОГО ОБУЧЕНИЯ «МЕШОК СЛОВ» В МОДУЛЕ «1С (ЧАТ-БОТ)» ДЛЯ АВТОМАТИЗАЦИИ РАБОТЫ СОТРУДНИКОВ ПЕРВОЙ ЛИНИИ ПОДДЕРЖКИ ПОЛЬЗОВАТЕЛЕЙ

УДК 331.101.52(004.85)

А.В. Зиновьев, ООО «Газпром информ» (Москва, Россия), a.zinovyev@inform.gazprom.ru

Г.Е. Язиков, ООО «Газпром информ», g.yazikov@inform.gazprom.ru

А.А. Аборнев, ООО «1С-КСУ» (Томск, Россия), a.abornev@1c-prombez.ru

В статье предложено описание практической реализации метода машинного обучения «мешок слов» (англ. bag of words), которая позволяет автоматизировать работу сотрудников первой линии поддержки, а также повысить эффективность и качество обслуживания. Решение реализовано на базе платформы «1С:Предприятие 8» (ООО «1С») в виде отдельного модуля с применением механизма расширений и виртуального окружения среды Python. Разработанный модуль может быть интегрирован со сторонними/внешними сервисами машинного обучения, содержит механизмы подготовки и выгрузки обучающих данных, собственный расширяемый API, а также включает модель машинного обучения для работы с текстом, которую можно адаптировать для прикладных задач по классификации «сырого» текста.

КЛЮЧЕВЫЕ СЛОВА: 1С, МАШИННОЕ ОБУЧЕНИЕ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, BAG OF WORDS, АВТОМАТИЗАЦИЯ.

Дочерняя сервисная ИТ-компания ПАО «Газпром» ООО «Газпром информ» осуществляет комплексное обслуживание и сопровождение информационных систем (ИС) и ИТ-инфраструктуры. В рамках сопровождения ИС сотрудники первой линии поддержки выполняют обязанности по обработке обращений пользователей. Системы постоянно развиваются, внедряются новые решения, соответственно, поток запросов с каждым годом возрастает. Появляется потребность в улучшении качества обслуживания и повышении эффективности сотрудников первой линии. В связи с этим необходим инструмент, который позволит автоматизировать обработку входящих пользовательских обращений и даст возможность сотрудникам первой линии поддержки получить необходимый ответ по запросу, дополнительно расширив их знания в некоторой области, а пользователю – готовое решение без дополнительной консультации.

В последнее время для решения подобных задач стали очень популярны средства обработки запросов в виде чат-бота с использованием алгоритмов и методов машинного обучения и искусственного интеллекта (ИИ).

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

Разрабатываемый в настоящее время национальный стандарт РФ дает следующее определение ИИ: «моделируемая, искусственно воспроизводимая

интеллектуальная деятельность мышления человека» [1]. Для практической реализации данной технологии используются специализированные алгоритмы, известные под общим термином «машинное обучение».

Область применения ИИ довольно широка. Это задачи по анализу и прогнозированию временных рядов, выявлению в них аномалий; обработке текстов на естественном языке, а также изображений и видеопотока в режиме реального времени и т.д. Искусственный интеллект используется также при создании рекомендательных систем, например чат-ботов и различных помощников.

По мере значительного увеличения доступности вычислительных ресурсов интерес к ИИ возрастает, появляются новые успешные проекты его применения в бизнесе.

Эффект от использования ИИ достигается за счет решения массовых и трудоемких задач, а также выявления скрытых закономерностей в данных.

Машинное обучение (англ. machine learning) – отдельный класс методов, для которых характерно не прямое решение одной задачи, а обучение в процессе решения множества сходных [2]. Для разработки таких методов используются различные разделы математики: статистика, вычислительная математика, математическое программирование, теория вероятностей и т.д.

Перечисленные выше направления применения ИИ можно свести к следующим категориям задач:



- регрессии, когда выявляются закономерности в числовых данных;
- классификации, когда некоторый набор признаков нужно отнести к определенному классу;
- кластеризации, когда есть набор данных с заранее неизвестными классами и необходимо по некоторым признакам разделить его на конечное количество классов.

Для решения каждой из перечисленных категорий задач используются различные методы машинного обучения. В настоящее время очень популярны те, что базируются на модели нейронных связей мозга – нейронные сети и глубокое обучение, где применяются нейронные сети со множеством слоев. Различные алгоритмы машинного обучения требуют и разных подходов: можно выделить обучение с учителем, без учителя, с подкреплением. Негласный стандарт в разработке подобных алгоритмов – язык Python. На нем создано огромное количество библиотек, содержащих новые алгоритмы для решения самых различных прикладных задач.

В ООО «Газпром информ» для обработки естественного языка при анализе запросов пользователей сотрудниками первой линии поддержки была применена наглядная и популярная концепция «мешок слов» (англ. bag of words), которая предполагает выполнение следующего конвейера методов.

- Стемминг и токенизация: слова обучающего корпуса приводятся к единым корням и разделяются на униграммы/биграмы (токены);
- частотный скоринг по TF-IDF: определяется важность отдельных токенов в зависимости от частоты употребления в классе и обучающем корпусе;
- отбор признаков (и ограничение размерности): Kbest;
- многоклассовая классификация с использованием методов машинного обучения: в зависимости от настроек применяются алгоритмы искусственной нейронной сети или логистическая регрессия.

Модель обучается классификации текста на данных информационной базы. От их качества зависят

точность и адекватность получаемых результатов. В модели содержатся встроенные механизмы по нормализации и балансировке данных перед обучением.

Основные требования для обучающих данных:

- размерность выборки. Минимальное количество примеров текста для одного класса – 10, ниже этого порога класс исключается из процесса обучения;
- делимость классов. Не должно быть пересечений в классификации (однотипных текстов, отнесенных к разным классам). В этом случае уверенность модели в отнесении к каждому из классов значительно снизится, ответы могут быть неточными;
- сбалансированность. Количество примеров по каждому из классов балансируется моделью, но нужно учитывать, что в ходе этого процесса общий размер выборки для всех классов будет ограничен таковым для класса с минимальным количеством примеров;

– особенности классификации текста. Модель машинного обучения в результатах классификации возвращает вероятность, с которой выбран именно этот класс (уверенность в ответе). Если уверенность ниже определенного порога, результат отсекается. При вводе наименования учитывается контекст. Например, если встречается слово, используемое в разных классах, модель с большой вероятностью выдаст их набор, но уверенность для всех классов будет низкой. В дальнейшем при уточнении текста вариативность может быть значительно снижена.

Модель обладает значительной способностью к обобщению, т.е. наименования новых элементов могут заметно отличаться от примеров, на которых производилось обучение. При добавлении классов в информационную базу после накопления достаточного количества примеров необходимо повторно обучить модель для использования новых данных [3].

ЧАТ-БОТ И БАЗА ЗНАНИЙ С ПРИМЕНЕНИЕМ МЕХАНИЗМОВ ПЛАТФОРМЫ «1С:ПРЕДПРИЯТИЕ»

В качестве инструмента взаимодействия модели с пользователем в модуль интегрировано решение,

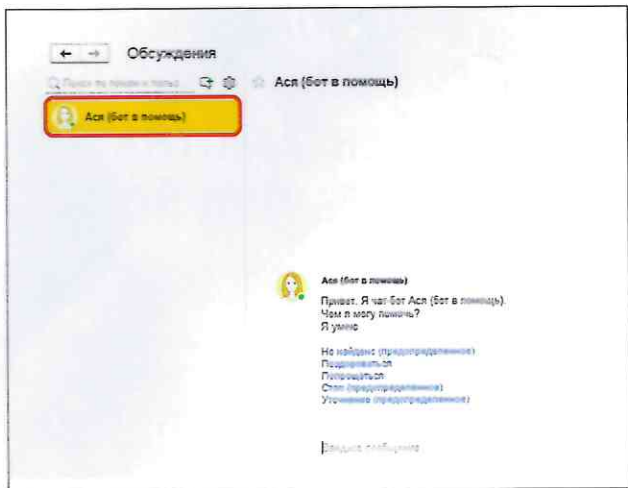


Рис. 1. Окно с чат-ботом «Ася (бот в помощь)»

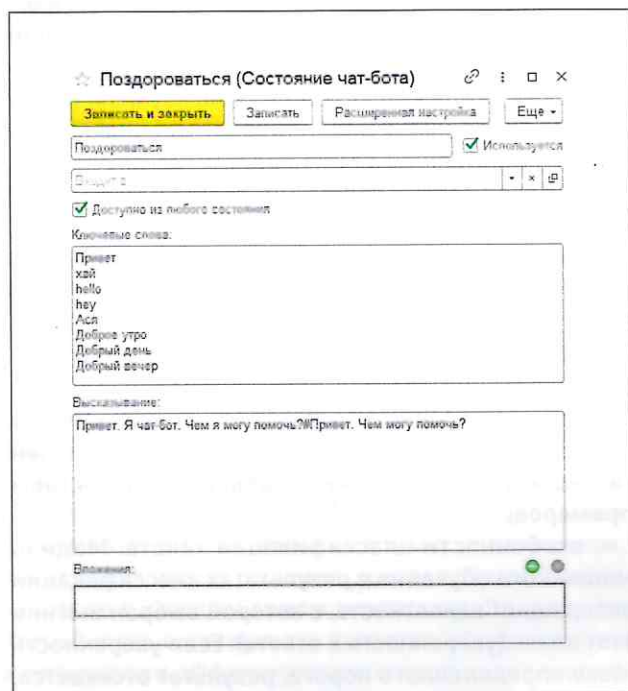


Рис. 2. Состояние чат-бота «Поздороваться»

разработанное ООО «1С», – чат-бот «Ася (бот в помощь)» (рис. 1), использующее механизмы платформы «1С: Диалог» [4].

Предварительно администратор создает в системе сценарии беседы чат-бота и устанавливает его состояние для различных ситуаций при общении с пользователем. Заранее необходимо добавить ключевые слова, по которым чат-бот будет искать это состояние, и фразу для ответа чат-бота, а также установить флаг использования. По такому принципу организованы, например, состояния «Поздороваться» (рис. 2) и «Попрощаться».

Доступна опция хранения истории бесед. При дальнейшем анализе это позволяет совершенствовать взаимодействие пользователя и чат-бота. Возможности

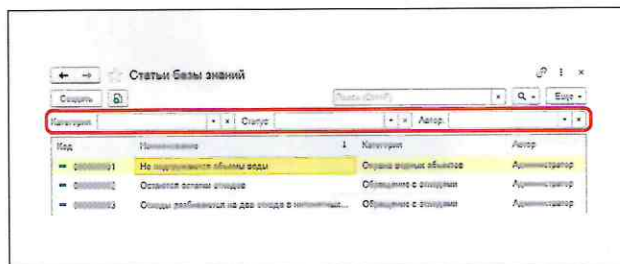


Рис. 3. Справочник «Статьи базы знаний»

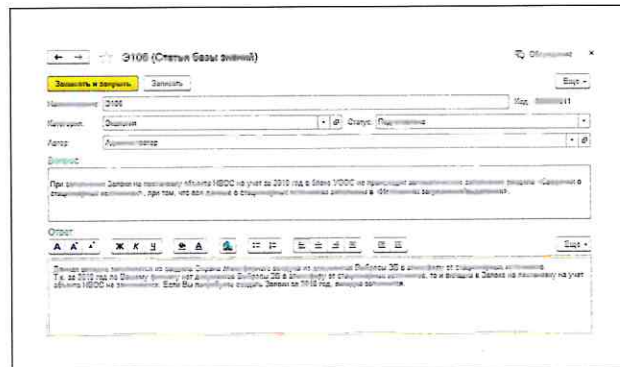


Рис. 4. Добавление новой статьи в базу знаний

последнего ограничены настройками, выполненными администратором. «Ася (бот в помощь)» определяет намерения по ключевым словам и фразам, содержащимся в тексте сообщения. Даже если пользователь допустит несколько ошибок в слове или сформулирует запрос образом, немного отличным от заложенных в настройках ключевых фраз, чат-бот с высокой долей вероятности поймет и ответит правильно, поскольку обладает некоторыми способностями к обобщению [5].

Использование чат-бота позволяет обратиться непосредственно к модели и получить результат обработки запроса, но для этого она должна быть предварительно обучена. Для достижения данной цели, а также дальнейшей обработки и совершенствования обучающей выборки был реализован объект модуля, позволяющий хранить последнюю в виде отдельных элементов системы, которые можно дополнять, изменять, а затем с их помощью выполнять переобучение. Существует справочник «Статьи базы знаний», предназначенный для накопления и управления знаниями на предприятии в рамках области данных прикладной информационной базы. Список статей предусматривает отборы по полям «Категория», «Статус», «Автор» [6].

Ввод новых статей может производиться как администратором (например, в соответствующей категории данных), так и пользователями через чат-бот «Ася (бот в помощь)» (рис. 4).

Подготовленный набор статей администратор может выгрузить и на их основе переобучить используемую модель.

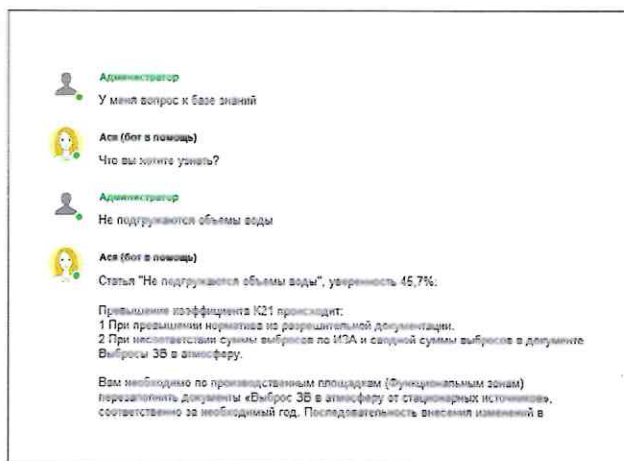


Рис. 5. Ответ и его точность (уверенность), выведенные чат-ботом



Рис. 6. Оценка качества статьи

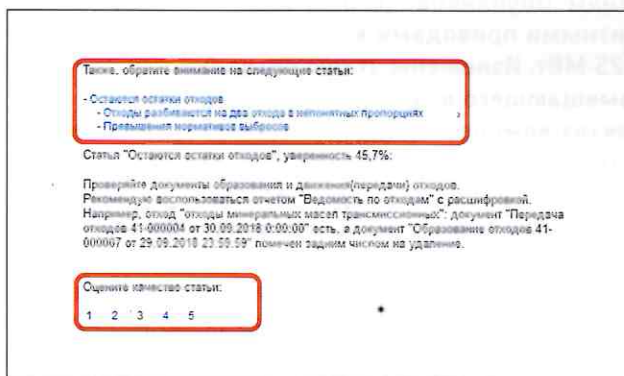


Рис. 7. Предложение других статей и оценка качества выбранной

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Пользователь платформы «1С:Предприятие» с подключенным модулем «1С (чат-бот)» (в рассматриваемом случае – сотрудник первой линии поддержки или пользователь) обращается к чат-боту. Последний

по заранее заложенным шаблонам состояний понимает, что пользователь задает вопрос для обработки ИИ, происходит обращение к модели обучения. Модуль производит подключение через созданный API (HTTP-сервисы и встроенные механизмы платформы «1С:Предприятие») к реализованной модели и отправляет запрос. Функции подключения заранее заданы в настройках чат-бота администратором, а их вызов происходит при переходе чат-бота в состояние ответа на вопрос пользователя. Обученная модель подбирает наиболее подходящие варианты решения. После обработки она отправляет на платформу «1С:Предприятие» наиболее подходящий ответ (ответы), и чат-бот выводит результат, указывая его точность (рис. 5).

В конце текста статьи сервис предлагает оценить результат подбора или (при отсутствии подходящего) добавить вопрос пользователя в «Базу знаний» для последующего ответа специалиста по знаниям (рис. 6).

Если есть такая необходимость, то можно выбрать другую статью. При этом чат-бот отобразит ее в обсуждении и даст возможность поставить оценку (рис. 7).

Наличие хорошо подготовленной обучающей выборки позволяет пользователю получить необходимый ответ и продолжить работу без привлечения дополнительных сотрудников или поиска информации в других источниках.

ВЫВОД

Разработанный модуль, включающий в себя решение на платформе «1С:Предприятие», модель машинного обучения, реализованную на Python, и средства интеграции, позволяет:

- повысить эффективность и качество обслуживания на первой линии поддержки;
- облегчить поиск и доступ к базе знаний по различным сферам деятельности компании;
- усовершенствовать предложенное решение путем оптимизации работы модели и механизмов взаимодействия чат-бота с пользователем.

В будущем планируется доработать функциональность модуля, обеспечить более точную работу модели и усовершенствовать интерфейсную составляющую взаимодействия с чат-ботом. ■

ЛИТЕРАТУРА

1. ГОСТ Р 43 0.8–2017. Информационное обеспечение техники и операторской деятельности. Искусственно-интеллектуализированное человекоинформационное взаимодействие. Общие положения [Электронный ресурс]. Режим доступа: <http://docs.cntd.ru/document/1200146327> (дата обращения: 17.02.2021).
2. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных: пер. с англ. М.: ДМК Пресс, 2015.
3. Григорович А., Афонин А., Бекетов В. 1С:Предприятие 8. Расширение «1С:Библиотека функциональных подсистем 1С-Совместно, редакция 3.0. Технологии искусственного интеллекта». Редакция 1.0. Руководство пользователя. М.: 1С-СОФТ, 2021.
4. Гацков А. Ася – бот в помощь: презентация: материалы Форума пользователей «1С:Документооборот», Москва, 17 апреля 2020 г. [Электронный ресурс]. Режим доступа: <https://v8.1c.ru/upload/iblock/b53/b53fe2218de4ef7c74fc325835a953fc.pdf> (дата обращения: 17.01.2021).
5. Григорович А., Афонин А., Бекетов В. 1С:Предприятие 8. Расширение «1С:Библиотека функциональных подсистем 1С-Совместно, редакция 3.0. Чат-бот ДО». Редакция 1.0. Руководство пользователя. М.: 1С-СОФТ, 2021.
6. Григорович А., Афонин А., Бекетов В. 1С:Предприятие 8. Расширение «1С:Библиотека функциональных подсистем 1С-Совместно, редакция 3.0. База знаний». Редакция 1.0. Руководство пользователя. М.: 1С-СОФТ, 2021.