

Библиотека TopicNet

Мощный набор инструментов на основе технологий ML для автоматизированного анализа больших слабоструктурированных текстов и мультимодальных данных

github.com/machine-intelligence-laboratory/TopicNet

Решаемые задачи

- оптимизация поиска в корпоративных базах знаний с учетом отраслевой специфики и терминологии;
- предопределение намерений потребителей по транзакционной активности;
- создание рекомендательных сервисов (экспертных и клиентских);
- анализ содержания новостных потоков

Отрасли

- Банки
- Телеком
- Медицина

Области применения

- Маркетинг
- Наука, образование, исследования
- Автоматизация бизнес-процессов (документооборот, call-центры)

Пример использования



Профилирование клиентов банка для последующего анализа групп потребления и персонализации предложений

Цель

Повысить конверсию продаж услуг существующим клиентам банка за счет улучшения персонализации предложений.

Проблема

Применяемая модель сегментации клиентов имела ряд недостатков и была малоэффективной: большая доля профилей не поддавалась интерпретации, а остальные профили сегментировались неверно. Как следствие, клиенты не интересовались дополнительными предложениями банка и жаловались на спам.




Результат

В результате проекта удалось снизить долю **неинтерпретируемых профилей до 0%** для клиентов с достаточными данными по транзакциям. Также был значительно **улучшен набор** интерпретируемых типов потребления клиентов банка. Для дальнейшего администрирования рекомендательной системы была **разработана модель**, которая раскладывает историю трат пользователя во времени и определяет его профиль. Данное внедрение позволило **повысить upsell-конверсию и лояльность клиентов**.

Суть решения

С помощью инструментов TopicNet мы проанализировали и сгруппировали данные транзакций клиентов на основе MCC-кодов по типам покупок. Кроме этого была разработана специальная модель, с помощью которой анализировались время и периодичность совершения покупок. Путем соотнесения с типовыми сценариями клиентского потребления был сформирован расширенный набор профилей, отражающий индивидуальный тип потребления каждого из клиентов.

Внедрение и лицензирование продукта

Исследования решаемости задачи на серверных мощностях МФТИ	Введение в опытную эксплуатацию	Перевод в промышленную эксплуатацию	Сопровождение
<ul style="list-style-type: none"> - анализ эффективности инструментов TopicNet и других (при необходимости); - функциональные требования к модулю в проде, включая требования к архитектуре; - кластера на основе обучающего множества или одной базы знаний; - дорожная карта проекта. 	<ul style="list-style-type: none"> - поставка лицензий модулей по прайсу; - внедрение в контур заказчика; - настройка взаимодействия с корпоративными источниками данных и существующими рекомендательными системами; - дополнительная разработка. 	<ul style="list-style-type: none"> - настройка инструментов администрирования; - обучение специалистов заказчиков для поддержки 1-й линии; - дополнительные настройки и отладки; - подготовка документации. 	<ul style="list-style-type: none"> - 2-я линия в режиме работы 5*8 Мск для профилактического и инцидентного дообучения модели (2-12 раз в год по ставке специалистов); - абонемент обновления ПО: бесплатно в течение первого года владения и 25% от стоимости ПО со второго года владения включительно.
 2-6 месяцев работы команды TopicNet	 от 3 месяцев работы команды TopicNet	 от 1 месяца работы команды TopicNet	

Партнеры:

ПАО Ростелеком

ПАО Сбербанк

ООО «1С»



Лаборатория машинного интеллекта



- 2 доктора физ.-мат. наук
- 12 специалистов Data Science Researcher & Engineer
- 4 внешних консультанта по исследованиям и разработкам

Возглавляет лабораторию **Константин Воронцов** (на фото – третий слева) – доктор физико-математических наук, профессор РАН, преподаватель МФТИ, МГУ, ВШЭ, ШАД Яндекс. Изобретатель в области алгоритмов ИИ, с помощью которых были созданы сложнейшие инструменты разведочного поиска, текстовой обработки.

Ключевые компетенции

1) Исследовательские AI-проекты

Разработанный на основе опыта лаборатории подход к реализации средне- и долгосрочных R&D проектов позволяет быстро провести полный цикл исследования и предоставить заказчику не только результаты анализа области, но и программную реализацию предлагаемого решения.

2) Основные направления исследований:

- Обработка данных носимых устройств (IMU-датчики)
- Fundamental AI: снижение вычислительной сложности модели
- TopicNet: анализ неразмеченных или слабоструктурированных данных



РАЗРАБОТКИ И ПРОГРАММНЫЕ РЕШЕНИЯ

- Открытая библиотека для тематического моделирования: BigARTM
- Библиотека автоматизированного тематического моделирования TopicNet
- Библиотека обработки изображений печатных документов
- Библиотека предобработки текстов

ТЕХНОЛОГИИ

- NLP: Topic Modeling, Knowledge Graph, Embeddings; выводы на основе текстовых коллекций
- CV: Object Detection & Segmentation, Video Processing, Super Resolution
- Time Series Processing: Wearable Devices, Predictive Maintenance

КОНТАКТЫ

CEO лаборатории машинного интеллекта,
руководитель проектов:

Гончаров Алексей
E-mail: alex.goncharov@phystech.edu
machine-intelligence.ru

Зам. директора по развитию
ЦК НТИ «Искусственный интеллект»:

Демидов Михаил
E-mail: demidov.ms@mipt.ru

ЦК НТИ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ» на базе МФТИ
141701, Московская обл., г. Долгопрудный
Научный пер., д. 4, к. 1 - МФТИ, корпус Физтех.Цифра
E-mail: ai@mipt.ru
Тел. +7 (498) 744-6617
ai.mipt.ru



Центр компетенций НТИ
«Искусственный интеллект»