

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В БРЕНДИНГЕ ТЕРРИТОРИЙ

Д. М. Концевой

*Белорусский государственный университет,
Минск, Беларусь, danielloysee@gmail.com
Научный руководитель – Ю. Ю. Гафарова,
кандидат философских наук, доцент*

В статье автор демонстрирует концептуальное применение актуальных практик из брендинга для построения стратегий развития территориальных брендов. В условиях трансформации роли ИКТ в обществе бренды становятся чрезвычайно подвижными символическими сущностями, а социальные сети – площадками интенсивного символического обмена. Формирование имиджа территорий во многом зависит от того, в какое русло направятся коммуникация и смыслообразование вокруг них. Для принятия оперативных решений необходимы инструменты сбора сигналов с подобных площадок и проектирования прогнозных моделей на их основе. Данная работа предлагает теоретический анализ современного территориального брендинга с практической демонстрацией создания и применения моделей машинного обучения для разработки и корректировки стратегий развития брендов территорий.

Ключевые слова: машинное обучение; брендинг территорий; брендинг городов; социальные сети; символический обмен; тематический анализ; анализ тональности; визуальный анализ.

В последние десятилетия понятие «бренд территории» пережило качественный переход от стабильного набора символов и официальных нарративов к непрерывно конструируемой сетевой практике. Традиционная модель, где имидж города или региона формировался преимущественно через институциональные кампании и медийные проекты, уступает место символической динамике, в которой ключевую роль играют информационно-коммуникационные технологии и платформы. Социальные сети и цифровые медиа превратили процесс смыслообразования в поток событий и реакций, где значимость символов определяется скоростью их репликации, характером вовлеченности и контекстом.

С развитием социальных сетей и инструментов публикации сообщений на «весь мир» вся бренд-парадигма начала свою трансформацию. Бренды из «статичных знаков», которых выпускают на рынок с заранее заданной нужной интерпретацией, превратились в поля символических битв между разными агентами сети. Сейчас для бренда важно, в каком направлении пойдет коммуникация вокруг него, а это, в свою очередь, во многом определяет его успех. Сегодня бренды – динамически изменяющиеся символические сущности. Эта же трансформация произошла с брендингом территорий. Бренд территории – совокупность знаков, ассоциаций и нарративов, постоянно реконструируемых агентами в сети в ходе коммуникации. По мнению автора книги «За пределами брен-

динга: как новые ценности прозрачности и честности меняют мир брендов» Николаса Инда, «хотя бренд связан с физическим продуктом или услугой, сам по себе он нематериален. Это преобразующая идея, которая превращает материальное в нечто ценное (although a brand is related to a physical product or service it is itself immaterial. It is a transforming idea that converts the tangible into something of value)» [3, с. 3].

Такой подход согласуется с современными теориями территориального брендинга. Кейт Динни в своей книге «Брендинг территорий. Лучшие мировые практики» пишет, что «в разработку бренда города вовлечено множество стейкхолдеров» [1, с. 10]. И хотя автор подразумевает под стейкхолдерами наиболее влиятельные городские группы, например, бизнесменов или политиков, мы смело можем причислять к ним и любые заинтересованные группы, участвующие в символическом обмене вокруг бренда города.

В структуре стратегии территориального бренда Кейт Динни выделяет элемент коммуникации [1, с. 27]. В настоящее время бренд-коммуникации не представляют собой процесс передачи сообщений пассивной аудитории. Бренд-менеджеры больше не могут контролировать бренды на рынке, наоборот, они своими действиями должны направлять коммуникацию вокруг бренда. Если же бренд-менеджерам нужно принимать правильные решения в отношении динамической коммуникации, у них должна быть перед глазами вся картина происходящего в сети. Возникает практическая проблема: как фиксировать, структурировать и прогнозировать те пласты смыслов, которые раньше исследовались преимущественно качественными методами? Появление больших потоков пользовательского контента: текстов, изображений, метаданных и сетевых взаимодействий, – даёт, с одной стороны, новые возможности для анализа и прогнозирования, с другой же стороны, ставит задачу работы с высокой неоднородностью и шумом.

В этой статье я предложу концептуальное применение актуальных практик из бренд-менеджмента и бренд-репутации в сфере брендинга территорий и опишу инструменты и методы, в том числе программный код на языке Python, с помощью которых можно строить как парсеры (программы для сбора информации), так и модели машинного обучения, способные в разы ускорить прогнозирование и помочь в разработке стратегий развития брендов территорий.

Начнем с того, с чего начинается любая ML-модель, а именно с цели работы. Любой этап EDA (Exploratory Data Analysis) и любая модель начинаются с определения того, что именно нужно собрать и что получить в итоге. В данной статье я затрону два типа данных, с которыми можно работать при анализе брендов территорий: вербальные тексты и изображения. В первом случае мы можем анализировать тематические рас-

пределения и нарративы, тональность, временные ряды и изменение тем во времени, а модели, построенные на этих данных, позволят прогнозировать тематическое соответствие постов, возможную силу отклика на сообщения. Во втором – визуальное представление, доминантные цвета, интерпретацию изображений, что может дать представление о соответствии визуала и территорий и позволит принять решение о том, какой визуал лучше использовать для той или иной территории. К этим же данным можно добавлять и метаданные, например, лайки, репосты, геоданные, с помощью которых можно оценивать силу акторов в сети, зависимости от географического расположения.

Кроме того, следует дать общую поэтапную логику построения любой модели ML, которой будем придерживаться в данной статье. Любая наша программа будет называться Pipeline, или конвейер. В программировании пайплайны удобны тем, что данные преобразуются последовательно, идут от этапа к этапу, что предотвращает утечку данных при работе с ними и закрепляет логику преобразования, пока данные не будут отправлены в модель. В качестве концептуального пайплайна я предложу следующую структуру: 1) сбор данных → 2) EDA → 3) разделение данных на тренировочную и тестовую выборки → 4) предобработка и нормализация данных → 5) создание меток, или кодирование данных (labeling/encoding) → 6) токенизация данных (embeddings) → 7) обучение модели → 8) оценка модели. В пайплайн можно добавлять и другие этапы, например, feature engineering, когда мы находим новые переменные для повышения точности модели, создание ансамблевых моделей ML, когда данные проходят через разные модели для получения лучших результатов. Все зависит от целей и задач [2].

Начнем с этапа сбора данных. Для наших целей мы будем собирать данные в социальных сетях, таких как X, Instagram, ВКонтакте и других площадках, на которых происходит символический обмен. Современные фреймворки позволяют без особых проблем собирать данные (вербальные тексты, фото, видео, метаданные) с любых сайтов. Сегодня повсеместно применяется библиотека Playwright, написанная для языка Python. Мощный инструмент автоматизации браузера, разработанный Microsoft, позволяет контролировать браузеры через единый API. С помощью данной библиотеки можно писать парсеры, которые способны обходить актуальные системы защиты браузеров (капчи, кнопки и т. д.), автоматически выстраивать безопасные тайм-ауты (время срабатывания действий), поддерживать работу прокси и асинхронную работу.

Любой парсер на Playwright начинается со следующего блока кода:

```
with sync_playwright() as pw:  
# Подключаем web driver  
browser = pw.chromium.launch(headless=False)  
# Создаем окно браузера
```

```

context = browser.new_context(viewport={"width": 1366, "height": 700})
# Открываем новую страницу в окне браузера
page = context.new_page()
# Переходим по ссылке на сайт социальной сети
page.goto('ссылка на соц. сеть')

```

Бот будет самостоятельно открывать браузер, создавать страницу и переходить по ссылке на сайт социальной сети. Если параметр `headless` равен `False`, то будет создано окно браузера, что позволяет контролировать работу парсера при его проектировании, однако в финальной версии лучше делать `headless = True` для экономии вычислительных ресурсов системы.

Далее можно применять самые различные инструменты веб-парсинга. С помощью метода `.locator()` следует искать нужный блок HTML-кода и работать с ним, например, находить окно ввода для запросов, вводить нужный запрос и нажимать `Enter`, затем находить нужный блок HTML, откуда и будет собираться информация. Методы `.get_attribute()`, `.inner_text()`, `.text_content()` нужны для сбора информации со страницы. Все методы можно посмотреть в документации библиотеки.

Как только нужные данные будут собраны, их следует выгрузить в удобные для работы форматы, например, `csv`, `xlsx`. В любом случае, все эти данные будут сгружены в датафрейм с помощью библиотеки `Pandas`, лучше использовать формат `csv`, так как он занимает меньше места на носителе.

Перейдем к существующим библиотекам, с помощью которых можно строить различные модели для реализации поставленных целей в рамках брендинга территорий.

Анализ текстов. Для анализа текстовых данных, связанных с брендами территорий, используются методы обработки естественного языка (NLP-модели), позволяющие исследовать тематическую структуру, тональность и динамику дискурсов. На первом этапе применяется классический подход с использованием векторизации TF-IDF, который преобразует тексты в числовые признаки, проще говоря, словам даются их значения, или векторы. Далее можно реализовывать регрессионные модели или градиентные бустинги, например `LightGBM`.

Ниже представлен простой код пайплайна логистической регрессии для нашей задачи (при условии, что `X`, `y` переменные определены):

```

from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, shuffle=True)

```

```

pipeline = Pipeline([
    ('tfidf', TfidfVectorizer(
        max_features=20000,
        ngram_range=(1,2),
        min_df=5,
        strip_accents='unicode'
    )),
    ('clf', LogisticRegression(max_iter=1000))
])
# Обучение модели
pipeline.fit(X_train, y_train)
# Предсказание и оценка модели
y_pred = pipeline.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))

```

Данный конвейер при правильно размеченных данных может делать простейшие операции определения тональности текстов с высокой точностью.

Более глубокий анализ возможен с помощью языковых моделей нового поколения, например, SBERT, и мультязычных трансформеров из библиотеки Hugging Face, например, ruBERT. Данные модели позволяют получать сложные эмбединги предложений, в которых сохраняются смысловые связи между словами и контекстами. На их основе можно решать задачи классификации сообщений, анализа тональности и тематического моделирования.

Анализ изображений. Для анализа изображений применяется стандартная архитектура сверточных нейронных сетей, которые требуют куда больших вычислительных мощностей, нежели обычные регрессионные модели или бустинги, чаще всего на основе предобученных моделей ResNet или EfficientNet из библиотеки torchvision. Эти модели позволяют извлекать визуальные эмбединги, или векторные представления изображений, которые затем можно анализировать или подавать в классические модели.

Таким образом, применение методов машинного обучения в брендинге территорий позволяет выйти за рамки описательных подходов и перейти к аналитическому пониманию символических процессов. Все описанные методы лишь часть того, что можно использовать в настоящее время в построении стратегий и принятии оперативных решений. Это своего рода конструктор, позволяющий находить связи там, где раньше их и не подумали бы искать.

Библиографические ссылки

1. Динни К. Брендинг территорий. Лучшие мировые практики. М.: Манн, Иванов и Фербер, 2013.
2. Рыжков А. Как построить хороший пайплайн разработки ML-модели // Хабр. URL: <https://habr.com/ru/articles/920030> (дата обращения: 28.10.2025).
3. Ind N. A Brand of Enlightenment // Beyond Branding. L.: Kogan Page, 2003.