

Прогнозирование Трендов в Медиа-Ландшафте: Подход на Основе Временных Рядов

Никита Петрович Ивкин

Московский физико-технический институт

Курс: Автоматизация научных исследований
(практика, В. В. Стрижов)/Группа 874

Эксперт: О. Н. Петров

Консультант: Ш. Л. Фоменко

2021

Цель исследования

Основная цель данного исследования - разработка комплексного подхода, объединяющего методы прогнозирования временных рядов и тематического моделирования, для точного прогнозирования трендов в динамичном медиа-ландшафте.

Общая схема предложенного подхода

Ключевые этапы
включают:

- ▶ Предобработку и кластеризацию данных
- ▶ Оценку значимости трендов
- ▶ Прогнозирование временных рядов
- ▶ Объединение прогнозов
- ▶ Обнаружение аномалий и улучшение прогнозирования

Предложенный гибридный подход обеспечивает точное и детальное прогнозирование динамики макро-ландшафта

Постановка задачи

Задача заключается в разработке эффективного метода прогнозирования популярных тем и трендов в медиа-пространстве, учитывающего сложную временную динамику и тематическую структуру данных.

Ключевые элементы решения:

- ▶ Многоуровневая кластеризация тем с использованием эмбедингов и тематического моделирования
- ▶ Оценка значимости кластеров на основе их позиционирования в медиа-ландшафте
- ▶ Применение модели Пророка для прогнозирования временных рядов каждого кластера
- ▶ Объединение прогнозов кластеров для получения комплексного прогноза
- ▶ Исследование методов обнаружения аномалий для улучшения прогнозирования

Вычислительный эксперимент

Результаты прогнозирования на реальных данных:

- ▶ Модель Пророка продемонстрировала высокую точность прогнозирования для кластеров, связанных с американским футболом и политикой
- ▶ Для некоторых кластеров, характеризующихся резкими пиками и изменениями трендов, традиционные методы прогнозирования показали ограниченную эффективность
- ▶ Средняя ошибка прогноза (MAPE) составила 28%, что характеризует хорошее качество прогнозов в рамках поставленной задачи

Основные результаты

- ▶ Предложен гибридный подход, сочетающий методы прогнозирования временных рядов и тематического моделирования
- ▶ Разработан механизм оценки значимости трендов, повышающий точность прогнозирования
- ▶ Выявлены ограничения традиционных методов прогнозирования при наличии аномалий в данных
- ▶ Намечены пути дальнейшего развития, включая исследование алгоритмов обнаружения аномалий

Представленный подход демонстрирует высокую эффективность в прогнозировании динамики медиа-ландшафта и может быть применен в различных областях, таких как анализ научных публикаций, прогнозирование спроса на продукты и мониторинг социальных тенденций.