

РАЗРАБОТКА МЕТОДА ПОИСКА ПРОСТРАНСТВЕННО-ПРОТЯЖЕННЫХ СОБЫТИЙ, РАЗВИВАЮЩИХСЯ ВО ВРЕМЕНИ

Юфа С.О. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»),

Вишератин А.А. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – д.т.н., доцент, Насонов Денис Александрович

(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Мониторинг событий является важным инструментом, используемым такими организациями, как СМИ или правительство. С ростом популярности социальных сетей, они становятся хорошим источником информации для мониторинга событий, однако для извлечения этих событий их нужно вычленивать из общего потока неструктурированной информации. Эта задача, а именно задача мониторинга событий с использованием социальных сетей, решается в этой работе.

В наше время быстро растет популярность социальных сетей. Благодаря этому, из них можно получить большое количество разнообразной информации, такой как текст, фото, видео, геолокация. Вся эта информация может быть использована для предоставления информации о важных для пользователя событиях. Социальные сети реагируют на события из самых различных сфер жизни, такие как концерты, выставки, демонстрации или автокатастрофы. Поэтому анализ социальных сетей становится удобным инструментом для обнаружения событий. При этом, задача обнаружения событий с использованием данных из социальных сетей осложняется наличием в социальных сетях большого количества посторонней информации. Извлечение информации, действительно полезной для идентификации событий, требует анализа времени появления сообщения, его геолокации, содержания и другой информации.

Существующие методы решения проблемы используют различные методы кластеризации, которые при разделении целевой области не учитывают сами данные, что снижает эффективность обнаружения событий. Для решения этой проблемы, был предложен метод мониторинга событий, учитывающий пространственное и временное распределение данных. Разработанный подход основан на построении нормального состояния рассматриваемой области для всех комбинаций месяца, типа дня (рабочий день или выходные) и часа, чтобы учесть изменения нормального состояния в зависимости от времени. Нормальное состояние рассматривается в виде дерева квадрантов, в котором каждый лист охватывает часть рассматриваемой области. Для лучшего учета распределения данных была разработана улучшенная версия дерева квадрантов, а именно сверточное дерево квадрантов, которое при построении использует расположение данных в пространстве. Использование сверточного дерева квадрантов позволяет точно различать области высокой и низкой активности в социальных сетях и, таким образом, повысить эффективность алгоритма обнаружения событий. Предложенный алгоритм поиска событий состоит из поиска аномалий, который в свою очередь состоит в сравнении текущей активности в ячейке с средней активностью в ней, и идентификации события, которая основана на построении графа хештегов в аномальной ячейке и рассмотрении его связанных компонентов. Так же в алгоритме проверяются соседние по времени и пространству ячейки для определения событий, протяженных во времени или пространстве.

Для разносторонней оценки разработанного подхода был собран исторический набор данных из социальной сети Instagram. В качестве исследуемого географического объекта использовался город Нью-Йорк, по причине высокой активности в социальных сетях в этом

городе. Полученный набор данных содержит более 62 миллионов постов в 114 тысяч геолокаций. Это позволило создать нормальное состояние на весь год и исследовать эффективность разработанного решения в двух экспериментах. Эксперименты показали, что было найдено 77% существующих событий и 73% найденных событий действительно существовало, что не уступает и даже превышает результаты передовых методов в данной области.

Юфа С.О.(автор)

Подпись

Вишератин А.А.(автор)

Подпись

Насонов Д.А. (научный руководитель)

Подпись