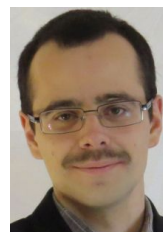


УДК 004.67

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ ДЛЯ РАННЕГО ОБНАРУЖЕНИЯ ЛЕСНЫХ ПОЖАРОВ



Д.Д. Дронов
магистрант гр. 155841,
кафедра ЭВМ, БГУИР



Д.Ю. Перцев
кандидат технических наук,
кафедра ЭВМ, БГУИР

Д.Д. Дронов

Является магистрантом Белорусского государственного университета информатики и радиоэлектроники по специальности 1-40 80 01 «Компьютерная инженерия (Хранение и обработка данных)».

Д.Ю. Перцев

Окончил аспирантуру Белорусского государственного университета информатики и радиоэлектроники (2016), защитил диссертацию в совете 05.13.01 (2020). Является доцентом кафедры ЭВМ.

Аннотация. Выявлены факторы для раннего обнаружения лесных пожаров, определены методы анализа и выполнен многофакторный анализ данных, представлена визуализация результатов анализа данных.

Ключевые слова: Интеллектуальный анализ данных, Многофакторный анализ, Регрессионный анализ, Визуализация многомерных данных, Метод главных компонент, Лесные пожары.

Введение

Лесные пожары – это серьезная угроза для окружающей среды и человеческой жизни. Они могут привести к уничтожению животных и растительных видов, земельных угодий и инфраструктуры. Пожары могут иметь серьезные последствия для экосистем лесов, угрожая их биоразнообразию и стабильности [1].

Необходимо принимать меры для предотвращения пожаров и контроля их распространения, а также обеспечивать эффективное и быстрое реагирование на возможные пожары для минимизации их негативных последствий.

В настоящее время для обнаружения лесных пожаров используются методы, пригодные для различных условий и территорий. Наиболее распространёнными методами являются:

- камеры видеонаблюдения. Они располагаются на высоте кромок деревьев, имеют обзор 360 градусов и способны обнаружить дым и его приблизительное местоположение. Возможны ложные обнаружения из-за естественных и техногенных испарений, которые похожи на дым;
- наблюдение со спутников. Получение снимков со спутников может занимать продолжительное время – от 2 до 12 часов;
- наблюдение с вышек и патрулирование;
- сообщения от граждан;
- беспилотные летательные аппараты.

Описание системы

Система раннего обнаружения лесных пожаров представляет собой множество устройств, объединённых в сеть ячеистой топологии [2][3] по протоколам IoT (Internet of Things). Каждое устройство содержит датчик температуры, углекислого газа, водорода и углеводородных газов [4][5][6].

В данной статье рассматривается метод многофакторного анализа данных для обработки информации от датчиков и других источников, которые собирают данные о пожаре в лесу. Все данные собираются в таблицу, пример которой представлен в таблице 1.

Входные данные включают дату и время, идентификатор устройства, геопозицию и показания с датчиков. Дополнительные факторы могут включать информацию о количестве осадков, силе и направлении ветра от метеорологических служб. Все эти данные обрабатываются с помощью многофакторного анализа данных.

Таблица 1. Пример данных, получаемых от устройств

#	ID	Дата	Время	GPS-широта	GPS-долгота	Температура	CO ₂	H ₂	C _x H _x
1	1	05.06.2022	15:00	29	52	28	0,0143	0,2906	0,9413
2	6	05.06.2022	15:01	34	47	28	0,0075	0,2603	1,4221
3	9	05.06.2022	15:02	37	44	28	0,0074	0,2139	0,9248
4	4	05.06.2022	15:03	32	49	28	0,0076	0,2579	1,9364
5	8	05.06.2022	15:04	36	45	28	0,0128	0,2607	0,1835
6	2	05.06.2022	15:05	30	51	28	0,0164	0,2237	0,0473
7	7	05.06.2022	15:06	35	46	28	0,0092	0,2095	0,3126
8	5	05.06.2022	15:07	33	48	28	0,0010	0,2297	0,6582
9	7	05.06.2022	15:08	35	46	28	0,0180	0,2136	1,8317
10	3	05.06.2022	15:09	31	50	28	0,0051	0,2329	1,4122

Многофакторный анализ данных

Метод многофакторного анализа данных является мощным инструментом, который позволяет определить взаимосвязи между несколькими переменными и определить значимые факторы, влияющие на исходную переменную.

Для данной системы анализ данных начинается с предварительной обработки входных данных, которые очищаются от выбросов, пропущенных значений и ошибок измерения. Далее выполняется нормализация данных, что позволяет сравнить различные переменные на основе их значений.

После этого строится модель многофакторного анализа данных. Данная модель включает в себя элементы, которые могут оказывать влияние на исходную переменную, – наличие пожара в лесу. Каждый фактор оценивается на основе коэффициента корреляции с исходной переменной [7].

Было обнаружено, что наиболее сильная связь наблюдается между температурой и содержанием углеродородных газов. Кроме того, было обнаружено, что сила и направление ветра существенно влияют на распространение пожара.

Результаты анализа показали, что данные можно разделить на несколько групп, основанных на геопозиции и количестве осадков. Таким образом, можно определить области, наиболее подверженные пожароопасности в зависимости от географического расположения и количества осадков.

Использование метода главных компонент (PCA)

Одним из методов многофакторного анализа данных является метод главных компонент (PCA). Он позволяет уменьшить размерность данных, выявить скрытые зависимости между факторами и сократить количество используемых переменных.

Представим, что мы имеем набор данных, включающий множество различных показателей, таких как температура, количество осадков, влажность, направление и скорость

ветра. Каждый показатель можно рассматривать как переменную. Такой набор данных может быть представлен в виде матрицы, где каждый столбец соответствует одному из показателей, а каждая строка - конкретному наблюдению.

Метод главных компонент позволяет определить, какие из показателей наиболее важны и объясняют наибольшую долю изменчивости в данных. Эти показатели называются главными компонентами. Остальные показатели могут быть отброшены или использованы в более узкой области исследования.

Таким образом, включение данных о метеорологических условиях в систему раннего обнаружения лесных пожаров может значительно улучшить ее точность и помочь предотвратить развитие пожаров в будущем.

Регрессионный анализ

Одним из методов анализа данных, который может быть применен для системы раннего обнаружения лесных пожаров, является регрессионный анализ, который позволяет определить связь между зависимой переменной и набором независимых, и использовать эту связь для прогнозирования значений.

В случае системы раннего обнаружения лесных пожаров, зависимой переменной может быть вероятность возникновения пожара, а независимыми переменными – температура, содержание углекислого газа, концентрация водорода и других газов, а также осадки, направление и скорость ветра. Для построения регрессионной модели необходимо использовать исторические данные о значениях этих переменных в периоды, когда происходили пожары.

Для построения регрессионной модели необходимо выбрать подходящий тип модели, определить зависимую и набор независимых переменных, а также оценить коэффициенты регрессии и проверить значимость полученных результатов.

После построения модели можно использовать ее для предсказания вероятности возникновения пожара на основе текущих значений переменных. Для этого необходимо ввести значения независимых переменных в модель и получить прогноз для зависимой. Если прогнозируемая вероятность возникновения пожара выше заданного порогового значения, система может автоматически сигнализировать об этом и предпринимать необходимые меры.

Регрессионный анализ может быть полезным инструментом для системы раннего обнаружения лесных пожаров, позволяя на основе исторических данных о пожарах и погодных условиях определить вероятность возникновения новых пожаров и своевременно предпринять меры для их предотвращения.

Визуализация данных

Для наглядного представления собранных данных используется геоинформационная система [8]. В зависимости от настройки параметров, возможно отображение датчиков и информации с них, либо раскрашивание карты в соответствии с риском возникновения пожара.

Лесной пожар вызывает отклонения значений датчиков. При этом различные виды пожара проявляются по-разному. Поэтому, при выявлении критериев важно опираться, по возможности, на данные с нескольких устройств, соседствующих друг с другом.

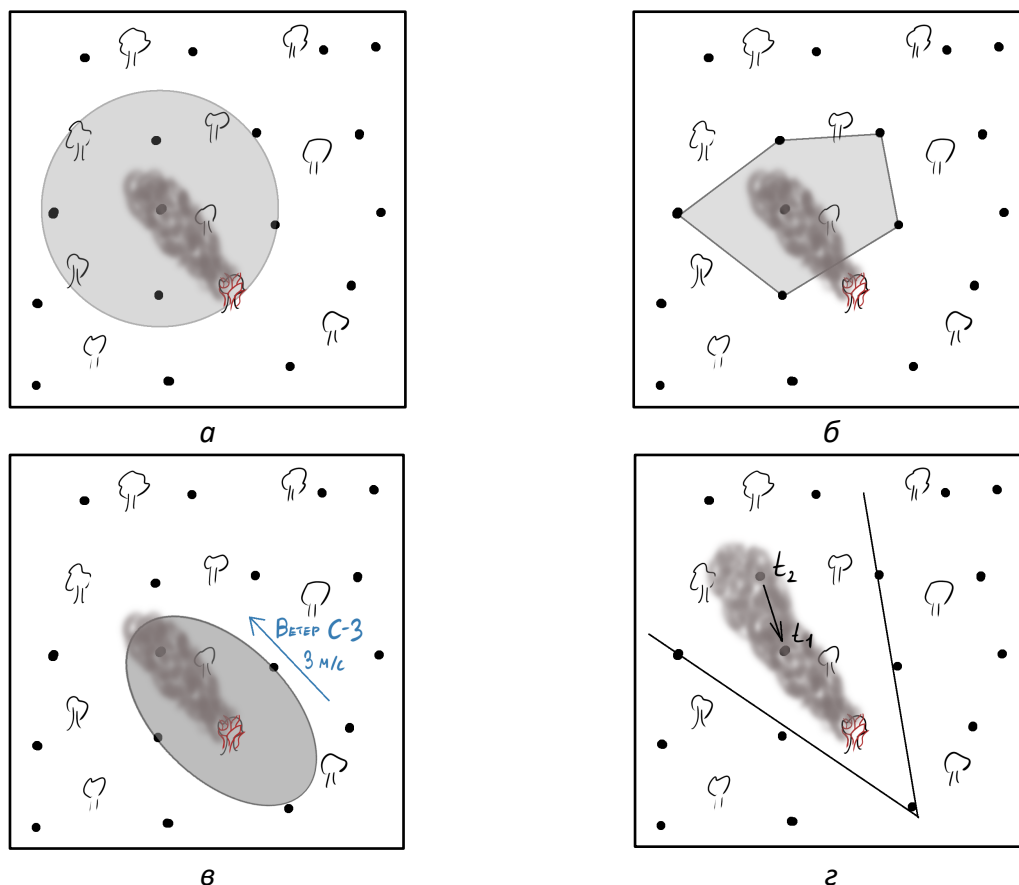


Рисунок 1. Определение области поражения: а - одно устройство; б – полигональное; в – с учётом метеоданных; г – на основе нескольких датчиков

Важно отметить, что в случае выявления признаков лесного пожара одним устройством, возможная область поражения описывается в общем случае радиусом, подходящим для мониторинга (рисунок 1.а), либо полигоном, границы которого определяются расстоянием между соседними датчиками (рисунок 1.б). Чтобы уточнить область поражения возможно использовать метеоданные, такие как скорость и направление ветра. В таком случае, возможная область поражения будет смещена в направлении, противоположном направлению ветра (рисунок 1.в). Если же один из датчиков обнаружил признаки пожара в момент времени t_1 , а второй – в момент времени t_2 , можно сделать вывод о направлении распространения продуктов горения (рисунок 1.г).

Пример анализа данных

В качестве примера анализа данных рассмотрим определение факторов, влияющих на вероятность возникновения лесного пожара в Слонимском районе Гродненской области Республики Беларусь. Для этого были собраны данные о погодных условиях, количестве осадков на территории лесного массива и возникновении пожаров. Данные были обработаны и проанализированы, чтобы выделить наиболее значимые факторы, влияющие на вероятность возникновения лесного пожара. В результате анализа данных методом PCA было выявлено, что наибольшее влияние на вероятность возникновения пожара оказывают три фактора: температура, концентрация углекислого газа и скорость ветра. Более высокие значения этих параметров соответствуют более высокой вероятности возникновения пожара. В результате анализа с помощью авторегрессионной модели ARIMAX [9] удалось построить модель риска возникновения лесного пожара, которая учитывает сезонные изменения исследуемых факторов.

Таким образом, были выделены ключевые факторы, влияющие на вероятность возникновения пожара в лесном массиве и построена модель для своевременного обнаружения.

Заключение

В данной работе была представлена система раннего обнаружения лесных пожаров на основе IoT устройств, объединенных в сеть ячеистой топологии. Были рассмотрены факторы, влияющие на пожароопасность, и выполнен анализ данных с помощью многофакторного анализа. Система раннего обнаружения лесных пожаров является важной мерой для предотвращения возникновения и быстрой локализации пожаров. Для ее эффективности необходимо учитывать множество различных факторов, таких как расположение устройств, типы датчиков и метеорологические условия.

Список литературы

- [1] Пожары меняют климат, даже когда мы об этом забываем [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://greenpeace.ru/blogs/2019/09/23/pozhary-menjajut-klimat-dazhe-kogda-my-ob-jetom-zabyvaem/> Дата доступа: 15.03.2023
- [2] Wireless Mesh Networks in IoT Networks / Yu Liu, Xiangdong Qiu [et al.] // 2017 International Workshop on Electromagnetics: Applications and Student Innovation Competition, 2017. – С.55-57.
- [3] Early Forest Fire Detection Using a Protocol for Energy-Efficient Clustering with Weighted-Based Optimization in Wireless Sensor Networks / Puneet Kaur [et al.] // Applied sciences, 2023. – <https://doi.org/10.3390/app13053048>.
- [4] Early forest fire detection using low-energy hydrogen sensors / K. Nörthemann [et al.] // Journal of Sensors and Sensor Systems, 2013. – С.171-177.
- [5] Wildland fire emissions, carbon, and climate: Emission factors / Shawn Urbanski // Forest Ecology and Management, 2014. – С.51-60.
- [6] Early forest fire detection and verification using optical smoke, gas and microwave sensors / Wolfgang Krüll [et al.] // Procedia Engineering, 2012. – С.584-594.
- [7] «Сырые» данные и некоторые рецепты их «приготовления» / М.М. Татур [и др.] // Информационные системы и технологии = Information Systems and Technologies : материалы междунар. науч. конгресса по информатике. В 3 ч. Ч. 1, 2022. – С.194-203.
- [8] Зиновьев, А.Ю. Визуализация многомерных данных / А.Ю. Зиновьев. – М. : Изд-во КГТУ, 2000. – 180 с.
- [9] Тихонов, Э.Е. Методы прогнозирования в условиях рынка: учебное пособие / Э.Е. Тихонов – Невинномысск, 2006. - 221 с.

DATA MINING FOR EARLY DETECTION OF FOREST FIRES

D. Dronov

*master's student of group 155841,
Electronic Computing Machines Department,
BSUIR*

D. Pertsau

*PhD,
Electronic Computing Machines
Department, BSUIR*

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

E-mail: bsuir.dmitriy.dronov@gmail.com, pertsev@bsuir.by

Abstract. Factors for early detection of forest fires are identified, analysis methods are determined and multivariate data analysis is performed, visualization of data analysis results is presented.

Key words: Data Mining, Multifactor analysis, Regression analysis, Visualization of multidimensional data, Principal component analysis, Forest fires.