*УДК 618…*

**ПОИСК ЗАКОНОМЕРНОСТЕЙ В ОЦЕНКЕ ЛЕСОПОЖАРНОЙ ОБСТАНОВКИ**

**ПО ПОГОДНЫМ УСЛОВИЯМ**

*М.И. Костенчук*

*младший научный сотрудник Н.А. Дрожжин*

*научный сотрудник Белоусов Р.Л.*

*ФГБОУ ВПО «Академия гражданской защиты МЧС России»*

***Аннотация.*** *В статье рассмотрен подход к оценке лесопожарной обстановки, в основу которого положен поиск ассоциативных связей между погодными условиями и лесными пожарами. Для построения ассоциативных правил проводится предобработка исходных данных с помощью алгоритмов кластеризации.*

*Ключевые слова: интеллектуальный анализ данных, кластеризация, метод k-средних, алгоритм APRIORI, временные ряды, нечеткие множества, нечеткая продукционная система.*

**SEARCH OF PATTERNS IN ESTIMATION OF FOREST FIRE SITUATION**

**BY WEATHER CONDITIONS**

*M Kostenchuk*

*junior researcher N Drozhzhin*

*research associate R Belousov*

*Civil Defence Academy EMERCOM of Russia*

***Abstract.*** *The article describes the approach to the assessment of forest fire situation, which is based on search of associative links between weather conditions and forest fires. For constructing association rules is conducted pretreatment baseline data using clustering algorithms.*

*Keywords: Data Mining, clustering, k-means method, Apriori method, time series, fuzzy sets, fuzzy production system.*

**Введение**

На территории России ежегодно регистрируется от 15 до 40 тысяч лесных пожаров, охватывающих площади до 2,5 млн га [6]. Ежегодный ущерб, который наносят пожары, исчисляется в миллиардах рублей. Например, в 2013 году ущерб составил порядка 20 миллиардов [7].

Из анализа данных службы государственной статистики (рис. 1) следует, что количество пожаров на территории России за последние 20 лет не имеет тенденции к уменьшению [6].

|  |  |
| --- | --- |
| *а)* | *б)* |

*Рисунок 1. Лесные пожары в Российской Федерации (на 1 ноября 2013г.):*

*количество пожаров (а), площадь пожаров (б)*

С целью предупреждения природных лесных пожаров целесообразно делать прогноз пожароопасной обстановки, который заключается в оценке количества и площади лесных пожаров. Опираясь на выводы, которые были сделаны в [2], можно заключить, что оценка пожароопасной обстановки в большей степени зависит от погодных условий: температуры, влажности воздуха, осадков и других. В статье предложен подход автоматизации поиска закономерностей между лесными пожарами и погодными условиями.

**Исходные данные**

Исходные данные для анализа представлены обобщенными значениями погодных условий и пожаров по неделям (табл. 1). В исходные данные также включены различные модификации погодных условий, например, изменение температуры и влажности за три, за две и за одну неделю. Эти параметры введены в рассмотрение для того, чтобы учесть динамику пожароопасных факторов.

Таблица 1

Данные по погодным условиям в Сковородинском районе

Амурской области **[5]**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Дата | Т, °C | ΔT1, °C | ΔT2, °C | ΔT3, °C | В, % | ΔВ1, % | ΔВ2, % | … | Пож. |
| 23.06.2008 - 29.06.2008 | 21,23 | 1,23 | 2,44 | 5,5 | 70,22 | -7,34 | 4,72 |  | 24 |
| 30.06.2008 - 06.07.2008 | 20,5 | -0,73 | 0,49 | 1,71 | 72,82 | 2,6 | -4,74 |  | 13 |
| 07.07.2008 - 13.07.2008 | 19,06 | -1,44 | -2,17 | -0,95 | 74,62 | 1,8 | 4,4 |  | 1 |
| 14.07.2008 - 20.07.2008 | 20,58 | 1,52 | 0,08 | -0,66 | 72,02 | -2,6 | -0,8 |  | 2 |
| … | … | … | … | … | … | … | … | … | … |

**Постановка задачи**

В [2] была рассмотрена математическая модель оценки лесопожарной обстановки, в основу которой положена нечеткая продукционная система [3]. Нечеткая продукционная система позволяет давать прогноз количества и площади лесных пожаров на определённый промежуток времени. Основным элементом этой системы являются нечеткие продукционные правила вида:

|  |  |  |
| --- | --- | --- |
|  | **ЕСЛИ** Температура *высокая* **И** Влажность *низкая*, **ТО** Кол-во пожаров *большое*  **ЕСЛИ** Температура *средняя* **И** ΔВ3 *сильно убывает*, **ТО** Кол-во пожаров *среднее* | (1) |

Такие правила позволяют судить об интенсивности природных лесных пожаров, т.е. представляют собой закономерности оценки лесопожарной обстановки по погодным условиям.

Научная задача заключается в автоматизации поиска закономерностей оценки лесопожарной обстановки по погодным условиям.

**Решение задачи**

*Перевод численных значений исходных данных в качественные*

Для установления зависимости между погодными условиями и лесными пожарами необходимо провести предварительную обработку данных. Предлагается разбить исходные данные на группы (кластеры) и каждой группе поставить в соответствие порядковый номер.

Группировку значений каждого признака погодных условий будем производить с помощью алгоритмов кластеризации [4], в частности, методом к-средних [1]. Процедура кластеризации позволяет упорядочить объекты в сравнительно однородные группы, например можно выделить кластер низких температур или кластер очень высоких температур для данного региона.

Результаты кластеризации исходных данных из табл. 1 представлены в табл. 2.

Таблица 2

Данные по погодным условиям разбитые по кластерам

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Дата** | Т | ΔT1 | ΔT2 | ΔT3 | В | ΔВ1 | ΔВ2 | ΔВ3 | Пожары |
| 24.03.2008 - 30.03.2008 | 5 | 3 | 4 | 4 | 4 | 2 | 4 | 3 | 4 |
| 31.03.2008 - 06.04.2008 | 5 | 3 | 3 | 3 | 4 | 3 | 2 | 4 | 3 |
| 07.04.2008 - 13.04.2008 | 5 | 3 | 2 | 3 | 4 | 3 | 4 | 3 | 1 |
| 14.04.2008 - 20.04.2008 | 5 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 1 |
| … | … | … | … | … | … | … | … | … | … |

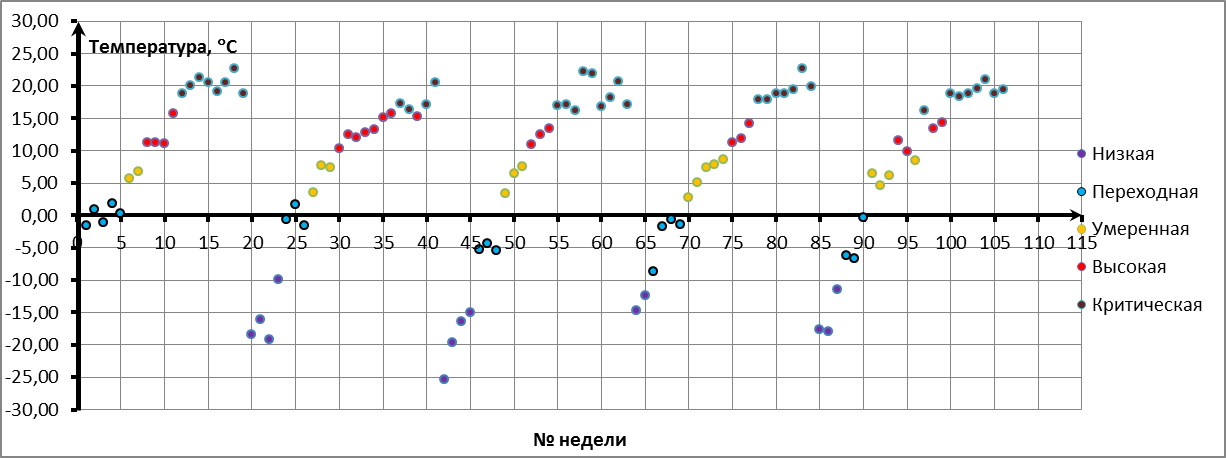
Основная идея метода к-средних заключается в том, что на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем множество разбивается на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Действие алгоритма таково, что он стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров:

|  |  |
| --- | --- |
|  | (2) |

где *k* – число кластеров (групп); *xj* – *j-ое* значение признака; *Si* – полученные кластеры;

*ai* – центры масс кластеров.

Ряд температур, разбитый на кластеры, представлен на рис. 2:



*Рисунок 2. Данные по температуре в пожароопасный период*

*с 2008 г. по 2012 г., разбитые на кластеры*

**Использование алгоритма APRIORI для определения часто повторяемых**

**погодных условий**

Для установления зависимости между погодными условиями и лесными пожарами необходимо разбить таблицу данных (табл. 2) на *n* таблиц, где *n* – количество кластеров множества пожаров. Т.е. погодные условия, соответствующие определённому кластеру множества пожаров, заносятся в отдельную таблицу. Например, для первого кластера пожаров таблица выглядит следующим образом:

Таблица 3

Погодные условия соответствующие первому кластеру пожаров

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Т | ΔT1 | ΔT2 | ΔT3 | В | ΔВ1 | ΔВ2 | ΔВ3 |
| 2 | 5 | 4 | 4 | 4 | 4 | 3 | 3 |
| 2 | 4 | 4 | 4 | 3 | 2 | 3 | 3 |
| 2 | 3 | 3 | 4 | 2 | 2 | 1 | 2 |
| 4 | 4 | 4 | 4 | 1 | 2 | 1 | 1 |

Далее для каждой таблицы необходимо решить задачу поиска частых наборов.

*Постановка задачи поиска частых наборов.*

Пусть *m –* число рассматриваемых показателей погодных условий, *n* – число кластеров одного показателя, *I = {i1, i2, i3, ...ik},* где *k* = *m\*n* – множество кластеров всех показателей погодных условий. Набором *R* называется набор *x1, x2,… xj,…,* где *xj∈I и 1 ≤ j ≤ m*. Частым набором *Rч* называется такой набор, поддержка *supp* которого (т.е. частота встречаемости набора в таблице) больше заданной *sup\_min.*

Задача состоит в том, чтобы найти все наборы *Rч* с поддержкой *supp ≥ supp\_min*.

*Решение задачи поиска частых наборов.*

Для нахождения частых наборов используется алгоритм интеллектуального анализа данных APRIORI [1]. Главным его достоинством является использование свойства антимонотонности, позволяющее не учитывать заведомо редкие наборы, т.е. наборы, поддержка которых ниже заданного порога. Так же алгоритм отличается простотой реализации.

Проиллюстрируем работу алгоритма на приведённом выше примере (табл. 3).

*Первый этап алгоритма APRIORI*

Первым этапом алгоритма APRIORI является построение дерева частых наборов (рис. 3).

Пусть минимальный порог *supp\_min* равен *0,2.*Сначала создаётся пустая корневая вершина и в её потомки записываются все возможные показатели погодных условий. В рассматриваемом примере у каждого погодного фактора по пять кластеров.

До тех пор, пока это возможно к каждой вершине добавляются все вершины, лежащие на том же уровне и правее. В процессе этого вычисляется значение поддержки набора (частота встречаемости), соответствующего вершине. Этот набор состоит из самой вершины и её пути до корня.

В рассматриваемом примере для первого кластера пожаров (табл. 3) три из четырёх значений температуры в таблице принадлежат второму кластеру, следовательно поддержка вершины *T=2* на первом уровне равна *0,75* (рис. 3). Набор *«T=2,В=2,* Δ*Т3=4»*, составленный из пути от корня до вершины *ΔT3=4*встречается в таблице один раз, значит поддержка этой вершины равна *0,25*. Если это значение меньше заданного порога, то рассматриваемая вершина отбрасывается, в силу свойства *антимонотонности* [1]. Оно гласит, что если встречаемость набора меньше заданной частоты, то наборы следующего уровня, в которых содержится данный набор, не рассматриваются в силу того, что частота их встречаемости не может увеличиваться.

*Второй этап алгоритма APRIORI*

Вторым этапом работы алгоритма является обход дерева кандидатов в глубину для составления наборов. На данном этапе в рассматриваемом дереве все пути от корня к вершине представляют собой часто встречающиеся наборы *Rч*, а остальные наборы из дерева исключены, потому что их поддержка меньше заданного порога. Следовательно, необходимо обойти все вершины, входящие в дерево и составить наборы (табл. 4).

Таблица 4

Часто встречающиеся наборы погодных условий

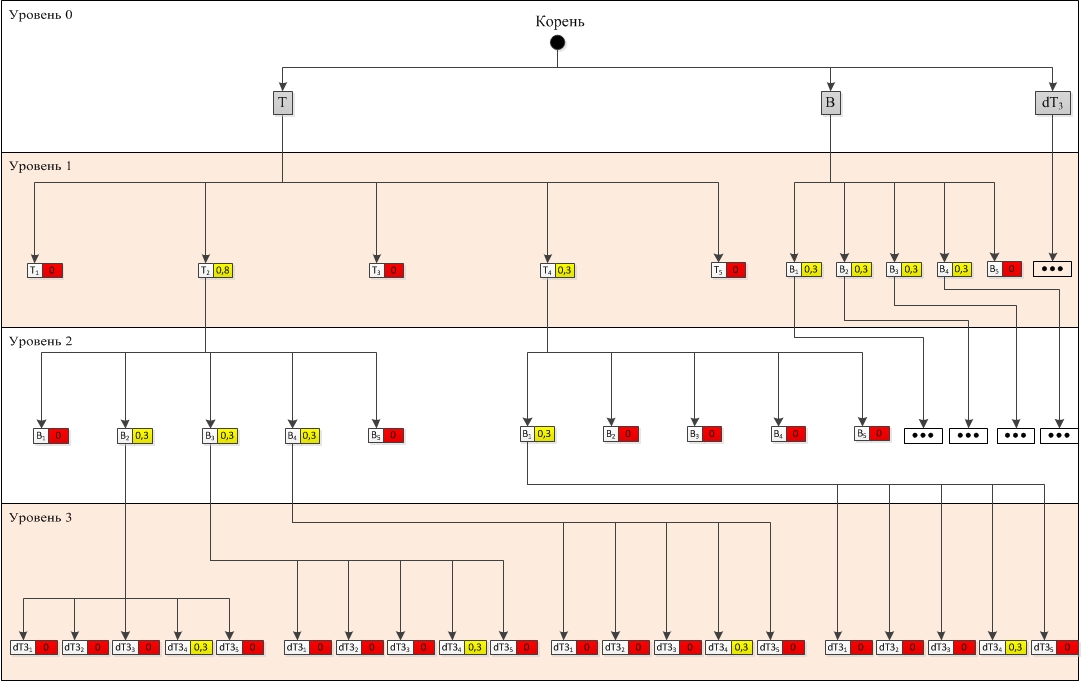
|  |  |
| --- | --- |
| № уровня | Наборы (поддержка) |
| 1 | *{Т=2 (0.75)}; {Т=4 (0.25)}; {В=1 (0.25)}; {В=2 (0.25)}; {В=3 (0.25)}; {В=4 (0.25)} и т.д.* |
| 2 | *{Т=2, В=2 (0.25)}; {Т=2, В=3 (0.25)}; {Т=2, В=4 (0.25)}; {Т=4, В=1 (0.25)} и т.д.* |
| 3 | *{Т=2, В=2, ΔТ3=4 (0.25)}; {Т=2, В=3, ΔТ3=4 (0.25)}; {Т=2, В=4, ΔТ3=4 (0.25)}; {Т=4, В=1, ΔТ3=4 (0.25)}* |

**Составление импликативных правил из найденных наборов**

Найденные часто повторяемые наборы (табл. 4) представляют собой наиболее характерные погодные условия для каждого из кластеров пожаров. Поэтому они являются условием в импликативных правил вида (1), а номер кластера пожаров является следствием правила. Иными словами в левую часть правила записываются наиболее свойственные определённому кластеру пожаров погодные условия, объединённые логической операцией «И». В правую часть правила записывается номер кластера пожаров, соответствующий табл. 3.

При составлении правил из наборов имеется ряд условий:

1. Ввиду того, что наиболее частые наборы, состоящие из одной погодной характеристики, не подходят для построения правил, а наборы, состоящие из максимального числа характеристик, являются очень редкими, возникает необходимость определить оптимальное количество переменных в наборе. Т.е. какие уровни дерева (рис. 3) рассматривать при построении правил.
2. Поскольку при составлении правил для каждого района будут учитываться различные погодные условия, проблематично установить минимальный уровень поддержки наборов. Поэтому необходимо определить количество наиболее весомых наборов (т.е. наборов с максимальной поддержкой), которые будут рассматриваться при составлении правил.

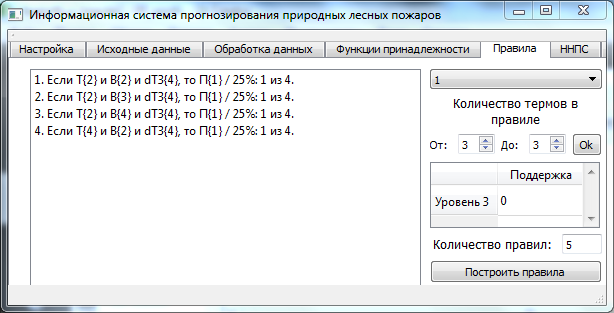


*Рисунок 3. Первый этап алгоритма Априори*

В настоящее время данные условия задаются экспертным путём. Для рассматриваемого примера при составлении правил берутся наборы третьего уровня (т.е. которые имеют три составляющие), и для одного кластера выбирается не более пяти правил с максимальной поддержкой (рис. 4).

**Вывод**

В статье был рассмотрен подход к автоматизации процесса составления правил для их использования в нечеткой продукционной системе. Процесс настройки правил имеет ряд условий, для решения которых требуется вмешательство эксперта, однако этот процесс можно автоматизировать. Нечеткая продукционная система с помощью правил сможет дать прогноз количества и площади лесных пожаров в районе на заданный промежуток времени.



*Рисунок 4. Составление правил (реализация в QT)*

**Список литературы:**

1. Барсегян А.А., Куприянов М.С., Степененко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – СПб.: БХВ-Петербург, 2004. – 336 с.: ил.
2. Н.А. Дрожжин, Р.Л. Белоусов, А.В. Воронецкий «Прогнозирование количества лесных пожаров по данным космического мониторинга»//научный журнал «Научные и образовательные проблемы гражданской защиты» №4 2013, С.11-15.
3. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTech. – СПб.: БХВ-Петербург, 2003. – 736 с.: ил.
4. Мандель И. Д. Кластерный анализ – М.: Финансы и статистика, 1988.—176 с: ил.
5. http://rp5.ru/Архив\_погоды\_в\_Игнашино [дата обращения: 03.03.2014].
6. http://www.gks.ru/free\_doc/new\_site/business/sx/les2.htm [дата обращения: 16.04.2014].
7. http://ria.ru/eco\_news/20131213/983924582.html [дата обращения: 16.04.2014]