***Белоусов Р.Л.***

***Дрожжин Н.А.***

***Костенчук М.И.***

**ВЫЯВЛЕНИЕ ЗАВИСИМОСТИ МЕЖДУ ПОГОДНЫМИ УСЛОВИЯМИ**

**И ЛЕСНЫМИ ПОЖАРАМИ**

*В статье рассмотрен один из возможных подходов определения зависимости между метеоусловиями и лесными пожарами, который основан на кластеризации и построении импликационных правил.*

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

***Belousov R.L.***

***Drozhzhin N.A.***

***Kostenchuk M.I.***

**RELATION BETWEEN WEATHER CONDITIONS AND FOREST FIRES**

*The article discusses one of the possible approaches determine the relationship between weather conditions and forest fires, which is based on clustering and building implicational rules.*

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

**Введение**

Россия по праву считается лесной державой, на неё приходится 1/5 часть всех лесов и половина всех хвойных лесов мира, леса занимают около 50% всей площади страны и составляют более миллиарда гектаров.

На территории России ежегодно регистрируется от 15 до 40 тысяч лесных пожаров, охватывающих площади до 2,5 миллионов гектар [gks], которые ежегодно наносят ущерб государству в миллиарды рублей. Например, в 2013 году ущерб составил порядка 20 миллиардов [ria].

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

|  |  |
| --- | --- |
| *а)* | *б)* |

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

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

Возможным способом, предупреждения пожароопасной обстановки и её последствий является прогнозирование количества и площади лесных пожаров.

Горимость лесов в большой степени зависит от погодных условий, например, таких как температура, влажность воздуха, скорость ветра и других [art2].

В данной статье рассматривается третий этап математической модели прогнозирования количества лесных пожаров [art2], а именно – определение зависимости между погодными условиями и лесными пожарами.

**Исходные данные**

Исходные данные для прогноза представляют собой набор временных рядов (табл. 1). В них входят усреднённые значения измерений в течение недели температуры, влажности, скорости ветра, других погодных условий и их модификации.

Таблица 1

**Исходные данные по погодным условиям в Сковородинском районе**

**Амурской области [rp5]**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Дата** | **Т, °** | **ΔT1, °** | **ΔT2, °** | **ΔT3, °** | **В, %** | **ΔВ1, %** | **ΔВ2, %** | **ΔВ3, %** | **…** | **Пож.** |
| 23.06.2008 - 29.06.2008 | 21,23 | 1,23 | 2,44 | 5,5 | 70,22 | -7,34 | 4,72 | -2,17 |  | 24 |
| 30.06.2008 - 06.07.2008 | 20,5 | -0,73 | 0,49 | 1,71 | 72,82 | 2,6 | -4,74 | 7,32 |  | 13 |
| 07.07.2008 - 13.07.2008 | 19,06 | -1,44 | -2,17 | -0,95 | 74,62 | 1,8 | 4,4 | -2,94 |  | 1 |
| 14.07.2008 - 20.07.2008 | 20,58 | 1,52 | 0,08 | -0,66 | 72,02 | -2,6 | -0,8 | 1,8 |  | 2 |
| … | … | … | … | … | … | … | … | … | … | … |

Модификации представляют собой изменения температуры и влажности за три, за две и за одну неделю. Эти параметры введены в рассмотрение для того, чтобы учесть тот факт, что, например, накопление влажности ведёт к увлажнению почвы и снижению возможности возгорания. Данным по погодным условиям ставится в соответствие количество пожаров, зарегистрированных с помощью космического мониторинга (табл.1).

**Постановка задачи**

В рассматриваемой математической модели прогнозирования лесных пожаров [art2] используется нечёткая продукционная система [fuzzy], которая позволяет спрогнозировать количество и площадь лесных пожаров на определённый промежуток времени. Элементами нечёткой продукционной системы являются импликационные правила вида:

|  |  |
| --- | --- |
| **ЕСЛИ** *Температура* высокая **И** *Влажность* низкая **ТО** *Кол-во пожаров* большое*;*  **ЕСЛИ** *Температура* средняя **И** *ΔВ 3* сильно убывает **ТО** *Кол-во пожаров* среднее*.* | (1) |

Такие правила обычно строятся экспертным путём, и их построение зависит от субъективной оценки эксперта. Необходимо решить задачу автоматизации процесса составления правил.

**Решение задачи**

1. **Перевод численных значений метеоусловий в качественные**

Для выявления зависимости между погодными условиями и лесными пожарами необходимо провести предварительную обработку данных. Для этого нужно некоторым образом группировать количественные значения погодных условий (табл.1) и заменить численные значения номером группы (табл. 2).

Таблица 2

Данные по погодным условиям разбитые по кластерам

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Дата** | **Т** | **dT1** | **dT2** | **dT3** | **В** | **dВ1** | **dВ2** | **dВ3** | **Пожары** |
| 24.03.2008 - 30.03.2008 | T5 | dT13 | dT24 | dT34 | В4 | dВ12 | dВ24 | dВ33 | П4 |
| 31.03.2008 - 06.04.2008 | T5 | dT13 | dT23 | dT33 | В4 | dВ13 | dВ22 | dВ34 | П3 |
| 07.04.2008 - 13.04.2008 | T5 | dT13 | dT22 | dT33 | В4 | dВ13 | dВ24 | dВ33 | П1 |
| 14.04.2008 - 20.04.2008 | T5 | dT13 | dT23 | dT33 | В4 | dВ13 | dВ23 | dВ33 | П1 |

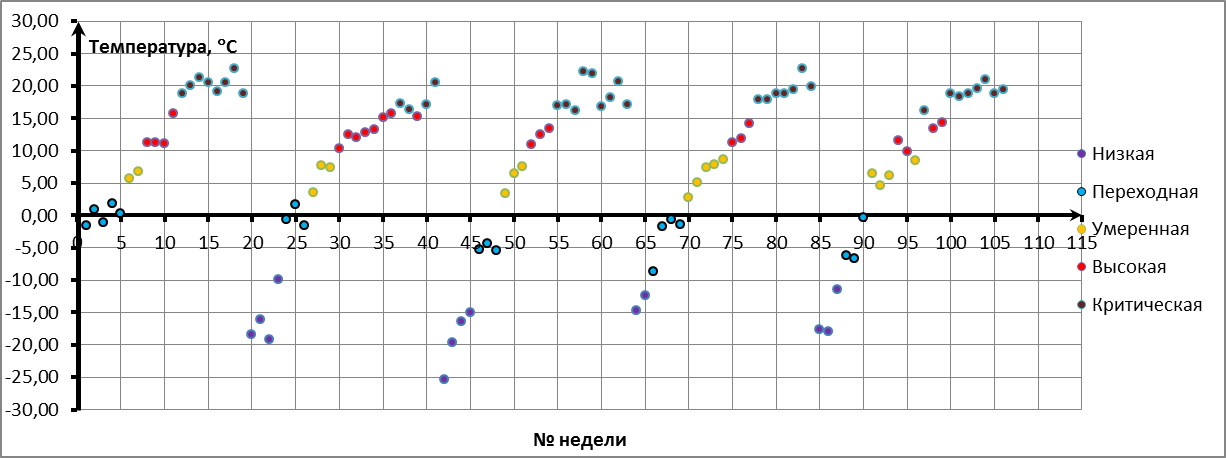
Группировка элементов каждого ряда метеоусловий производится с помощью кластеризации [mand]. Процедура кластеризации позволяет упорядочить объекты в сравнительно однородные группы. Кластеризация будет проводиться методом к-средних [barg]. Данный метод разбивает рассматриваемое множество на заранее заданное число кластеров. Например, множество температур разбивается на пять кластеров.

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; *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 по 2012 гг., разбитые на кластеры*

В результате кластеризации табл. 1 переводится в табл. 2.

1. **Использование алгоритма Априори для нахождения частых наборов погодных условий**

Для решения задачи выявления 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** | **dT1** | **dT2** | **dT3** | **В** | **dВ1** | **dВ2** | **dВ3** |
| T2 | dT15 | dT24 | dT34 | В4 | dВ14 | dВ23 | dВ33 |
| T2 | dT14 | dT24 | dT34 | В3 | dВ12 | dВ23 | dВ33 |
| T2 | dT13 | dT23 | dT34 | В2 | dВ12 | dВ21 | dВ32 |
| T4 | dT14 | dT24 | dT34 | В1 | dВ12 | dВ21 | dВ31 |

Далее для каждой таблицы не обходимо 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 *supp\_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 Априори. 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).

**2.1 Первый этап алгоритма Априори**

Первым этапом алгоритма Априори 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 это по пять термов температуры, влажности и изменения температуры за три недели.

Дальше, 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, соответствующего вершине. Этот набор состоит из самой вершины и её пути до корня.

Например, три из четырёх значений температуры в таблице принадлежат второму кластеру, следовательно поддержка вершины T2 на первом уровне равна *0,75*. Набор «T2В­2dТ4» встречается в таблице один раз, значит поддержка вершины dT3, из пути к которой составлен набор, равна *0,25*. Если это значение меньше заданного порога, то рассматриваемая вершина отбрасывается ввиду того, что все наборы, составленные из этой и любых других вершин, заведомо будут встречаться так же или менее часто в силу свойства *антимонотонности*. Оно гласит, что при добавлении к набору дополнительного элемента, его поддержка не может увеличиться.

**2.2 Второй этап алгоритма Априори**

Ввиду того, что наиболее частые наборы состоят из одной погодной характеристики и поэтому неинтересны для построения правил, а наборы, состоящие из максимального числа характеристик, являются очень редкими, возникает необходимость определить оптимальное количество переменных в наборе. На данном этапе нужное количество переменных в искомых наборах устанавливается экспертом вручную.

Вторым этапом работы алгоритма является обход дерева кандидатов в глубину для составления наборов. Необходимо перебрать все вершины, входящие в дерево на уровнях, соответствующих установленным значениям количества переменных в наборе. Каждый путь от корня к листу нужного уровня является часто встречающимся набором.

1. **Составление импликационных правил из найденных наборов**

Найденные наборы представляют собой наиболее характерные погодные условия для каждого из кластеров пожаров. Поэтому они используются как конъюнкт левой части правила. Иными словами в левую часть правила записываются наиболее свойственные определённому количеству пожаров метеоусловия, объединённые логической операцией «И». В правую же часть правила записывается кластер количества пожаров, соответствующий текущей таблице.

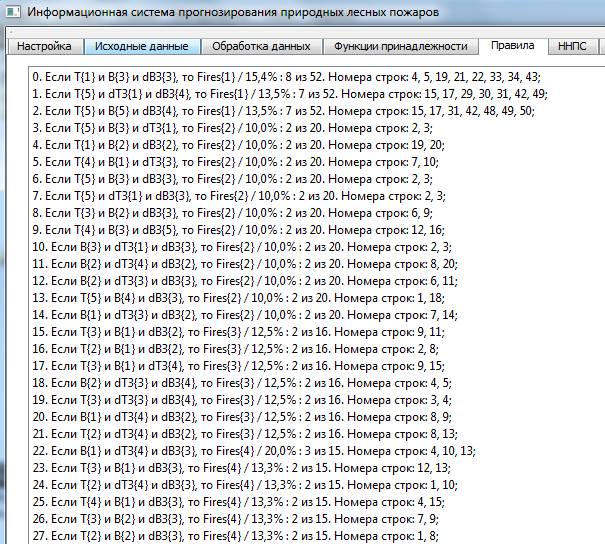


*Рисунок 3. Первый этап алгоритма Априори*

Из составленных правил выбирается несколько наиболее весомых в каждой из таблиц. Если правил с одинаковой поддержкой больше, чем установленное количество, то необходимо использовать их все, т.к. они обладают одинаковым весом. Например, для приведённого выше примера будут составлены следующие правила:

|  |  |
| --- | --- |
| **ЕСЛИ** *T = T2* **И** *В = B2* **И** dT3 *=* dT34 **ТО** *П = П1;*  **ЕСЛИ** *T = T2* **И** *В = B3* **И** dT3 *=* dT34 **ТО** *П = П1;*  **ЕСЛИ** *T = T2* **И** *В = B4* **И** dT3 *=* dT34 **ТО** *П = П1;*  **ЕСЛИ** *T = T4* **И** *В = B2* **И** dT3 *=* dT34 **ТО** *П = П1;*  *…* |  |

Описанный алгоритм был реализован на языке программирования C++ в среде Qt Creator. Для апробации были выбраны данные о пожарах и погодных условиях по Сковородинскому району Амурской области за пять лет с 2008 по 2012. Результат построения правил представлен на рис. 4.



*Рисунок 4. Скриншот найденных правил.*

**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]