***Дрожжин Н.А.***

***Белоусов Р.Л.***

**ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА ЛЕСНЫХ ПОЖАРОВ В РАЙОНЕ**

**И ОЦЕНКА КЛАССА ПОЖАРООПАСНОСТИ ТЕРРИТОРИИ**

*В статье рассмотрен один из возможных подходов прогнозирования количества лесных пожаров и оценки класса пожароопасности территории, который основан на кластеризации, построении ассоциативных правил, анализе временных рядов и создании нечеткой продукционной системы.*

*Ключевые слова:Data Mining, кластеризация, метод k-средних, ассоциативные правила, метод Априори, временные ряды, аппроксимация, нечеткие множества, нечеткая продукционная система, нечеткий логический вывод.*

***Drozhzhin N.A.***

***Belousov R.L.***

**PREDICTION OF THE NUMBER OF FOREST FIRES IN THE AREA**

**AND EVALUATION FIRE-RATED CLASS TERRITORY**

*The article discusses one of the possible approaches predict the amount of forest fires and fire-rated class territory, which is based on clustering, association rules, time series analysis and the creation of fuzzy production system.*

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

Лесные пожары наносят колоссальный ущерб экономике и экологии РФ, а также приводят к гибели людей. Учитывая масштабы и последствия природных пожаров на территории России можно сделать вывод о том, что прогнозирование количества лесных пожаров является актуальной задачей.

Проведенный анализ позволил выделить три вида математических моделей прогнозирования лесных пожаров:

* модели физико-химических процессов горения;
* модели прогнозирования контура одного пожара;
* моделей прогнозирования класса пожароопасности территории.

Модели третьего вида весьма немногочисленные и встречаются в основном в зарубежных источниках. Эти модели основаны на анализе данных статистики и выявлении тенденций возникновения лесных пожаров от определенных факторов.

В данной статье предлагается рассмотреть модель прогнозирования количества лесных пожаров (по количеству спрогнозированных пожаров можно сделать вывод о классе пожароопасности территории).

Возникает вопрос как измерить количество лесных пожаров. С 2009 г. МЧС России проводит космический мониторинг земной поверхности в том числе и на наличие термоточек. Термоточки – это тепловые аномалии, выявленные по результатам космической съемки в инфракрасном спектре. Данные по термоточкам обрабатываются с помощью автоматизированных алгоритмов и заносятся в базу данных по каждому району. Далее термоточки проверяются оперативными группами на местности и в базе помечается является ли термоточка пожаром.

Проведя анализ статистических данных термоточек была разработана математическая модель прогнозирования лесных пожаров (построение которой изложено в статье «Прогнозирование количества лесных пожаров по данным космического мониторинга»[1]).

Построение рассматриваемой математической модели состоит из 4-х этапов:

1. Установление наиболее значимых пожароопасных факторов, влияющих на лесные пожары (например температура окружающей среды, влажность воздуха).
2. Определение зависимости между выявленными факторами (на 1-ом этапе) и количеством лесных пожаров.
3. Прогноз значений каждого фактора на заданный интервал времени.
4. Прогноз количества лесных пожаров в районе на заданный интервал времени.

В данной математической модели на первом этапе определяются основные природные факторы, которые влияют на возникновение лесных пожаров. Этими факторами являются данные по погодным условиям. Статистические данные [2] по погодным условиям в районе представлены в табл. 1, где Т, В – средняя температура и влажность воздуха за неделю в районе, а ΔT 1, ΔT 2, ΔT 3, ΔВ 1, ΔВ 2, ΔВ 3 – прирост (падение) температуры или влажности за одну, две и три недели соответственно и др. Значение количества лесных пожаров представляется статистическим набором количества термоточек. Представленные данные характеризуют Сковородинский район Амурской области. Выбор Сковородинкого района обуславливается большими объемами лесных массивов, что снижает вероятность возникновения пожаров от антропогенного источника.

Таблица 1

Данные по погодным условиям в Сковородинском районе Амурской области

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Дата | Т | ΔT 1 | ΔT 2 | ΔT 3 | В | ΔВ 1 | ΔВ 2 | ΔВ 3 | … | Пожары |
| 24.03.2008 - 30.03.2008 | -1,35 | 6,26 | 6,35 | 6,49 | 69,75 | 8,17 | 5,15 | 0,46 |  | 13 |
| 31.03.2008 - 06.04.2008 | 0,98 | 2,32 | 8,58 | 8,67 | 62,43 | -7,32 | 0,85 | -2,17 |  | 15 |
| 07.04.2008 - 13.04.2008 | -0,97 | -1,94 | 0,38 | 6,64 | 55,51 | -6,92 | -14,24 | -6,07 |  | 20 |
| 14.04.2008 - 20.04.2008 | 1,87 | 2,84 | 0,90 | 3,22 | 49,26 | -6,25 | -13,17 | -20,49 |  | 20 |
| … | … | … | … | … | … | … | … | … | … | … |

На втором этапе построения модели вводятся нечеткие лингвистические переменные [3] (каждый столбец табл.1 это лингвистическая переменная) и каждому численному значению из табл. 1 ставится в соответствие имя терма соответствующей лингвистической переменной у которого значение функции принадлежности наибольшее (рис.1).

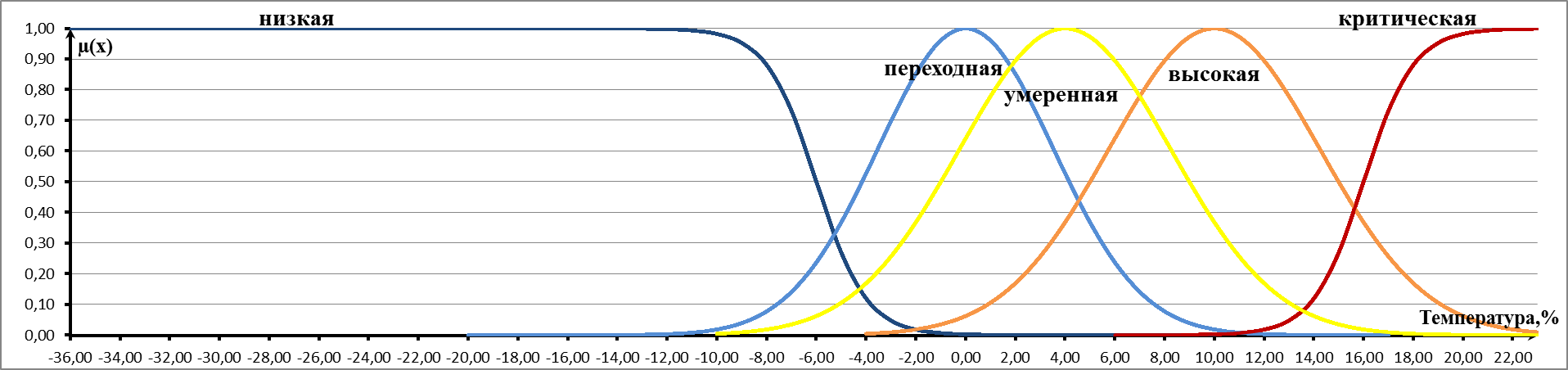


Рис.1 – Лингвистическая переменная температура воздуха

Преобразования проводятся для каждого столбца и в результате численные значения данных по метеоусловиям переводятся в нечеткие (табл.2).

Таблица 2

Данные по погодным условиям в лингвистическом виде

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Дата | Т | ΔT 1 | ΔT 2 | ΔT 3 | В | ΔВ 1 | ΔВ 2 | ΔВ 3 | Пожары |
| 24.03.2008 - 30.03.2008 | пере  ходная | сильно возрастает | сущест  венно возрастает | сущест  венно возрастает | высокая | сущест  венно возрастает | изменяется незначи  тельно | изменяется незначи  тельно | среднее |
| 31.03.2008 - 06.04.2008 | пере  ходная | сущест  венно возрастает | сущест  венно возрастает | сущест  венно возрастает | умерен  ная | сущест  венно  убывает | изменяется незначи  тельно | изменяется незначи  тельно | большое |
| 07.04.2008 - 13.04.2008 | пере  ходная | изменяется незначи  тельно | изменяется незначи  тельно | сущест  венно возрастает | понижен  ная | сущест  венно убывает | сущест  венно  убывает | убывает | большое |
| 14.04.2008 - 20.04.2008 | пере  ходная | сущест  венно возрастает | изменяется незначи  тельно | изменяется незначи  тельно | низкая | сущест  венно убывает | сущест  венно убывает | сущест  венно убывает | большое |

Перевод значений в нечеткий вывод обуславливается тем, что необходимо найти зависимость между погодными факторами и количеством пожаров. В случае с нечеткой классификацией значений погодных условий необходимо выделить 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) |

На третьем этапе построения модели осуществляется прогноз погодных факторов на заданный интервал времени путем аппроксимации соответствующих данных рядом Фурье вида:

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

Аппроксимация ряда температуры представлена на (рис.2).

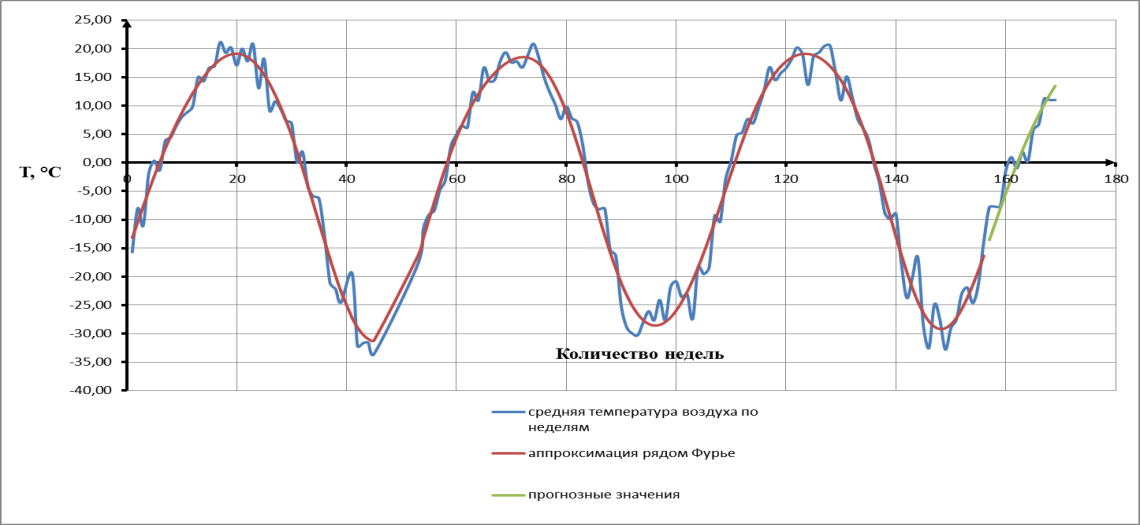


Рис. 2 – Аппроксимация температурных данных за 2005-2007 гг.

рядом Фурье и прогноз

На четвертом этапе построения модели строится нечеткая продукционная система (НПС) [3]. которая представляет собой прогностическую модель, основанную на правилах логического вывода. Правила вида (1) входят в базу правил НПС. На вход системы поступают погодные условия за прошедшую неделю, а также за три недели – система выдает прогнозное значение количества лесных пожаров на следующей неделе. Если на вход НПС поступают спрогнозированные значения природных условий на *n*-ую неделю вперед, нечеткая продукционная система выдает число пожаров на *n+1*-ую неделю. НПС позволяет сделать прогноз количества лесных пожаров на 8 недель вперед.

Построение рассмотренной математической модели имеет ряд «узких мест»:

* правила вида (1) строятся вручную путем глубокого анализа статистических данных (понадобится большое количество экспертов ввиду большого объема данных и их индивидуальности для каждого района);
* лингвистические переменные («Температура», «Влажность» и др.) строятся экспертным путем и зависит от ощущений эксперта (присутствует субъективизм, также большой объем данных);
* численные значения переводятся в нечеткие, привязываясь к построению термов ЛП.

Для учета вышеизложенных недостатков предлагается:

1. Автоматизировать процесс построения лингвистических переменных и перевод численных значений в нечеткие;
2. Автоматизировать процесс построения базы правил нечеткой продукционной системы, основываясь на статистических данных.
3. ***Автоматизация построения лингвистических переменных***

На рис. 3 представлено четкое множество температуры окружающей среды за пожароопасный период (т.е. рассмотрены только весна и лето). Проведем кластеризацию [4] этого множества. Процедура кластеризации позволяет упорядочить объекты в сравнительно однородные группы. Воспользуемся методом k-средних [4], чтобы разбить рассматриваемое множество на кластеры. Заранее определим, что кластеров будет пять.

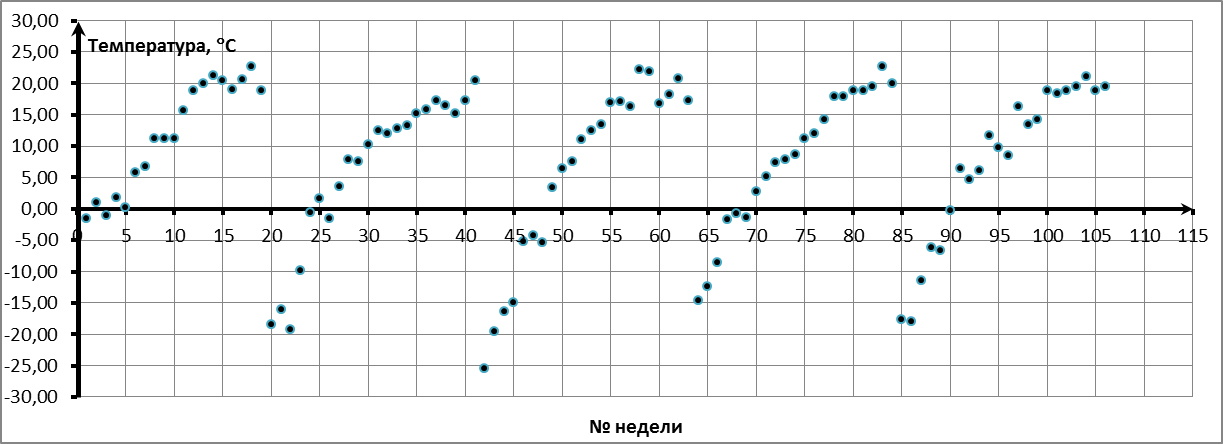


Рис. 3 – Температурных данные за 2008-2012 гг.

в пожароопасный период

Основная идея метода k-средних заключается в том, что на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Начальные центры масс выбираются случайно. Действие алгоритма таково, что он стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров (3):

|  |  |
| --- | --- |
|  | (3) |

где *k* – число кластеров; *Si* – полученные кластеры; *ai* – центры масс кластеров

Ряд температур, разбитый на кластеры представлен на рис.4.

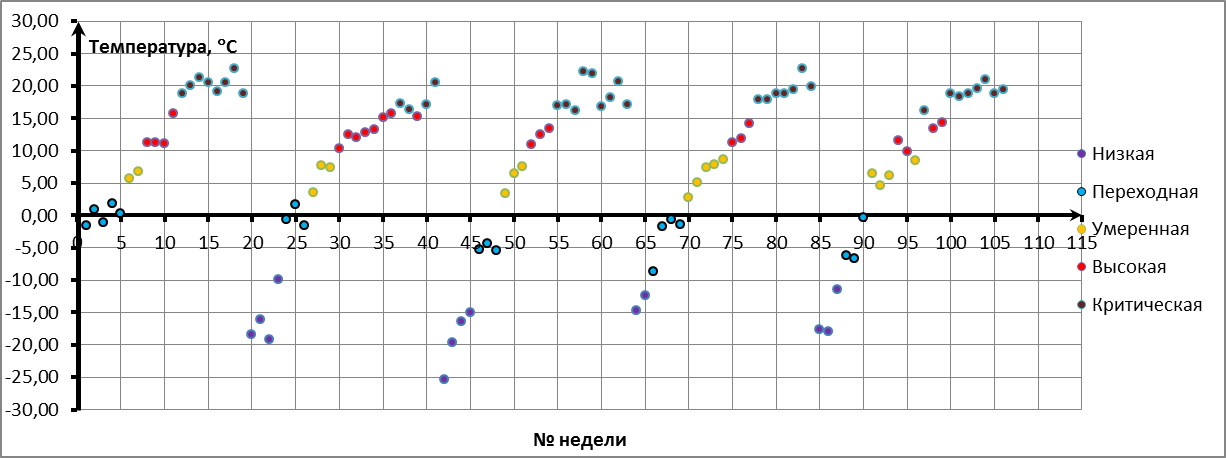


Рис. 4 – Температурных данные за 2008-2012 гг.

разбитые по кластерам

Каждому кластеру имеет название схожее с названием соответствующего терма лингвистической переменной «Температура окружающей среды», 1-й – «Низкая», 2-й – «Переходная» и т.д. Будем говорить, что если отдельно выбранное значение температуры попадает в i-й кластер, то в табл. 2 записываем имя этого кластера (т.е. в отличии от рассмотренного выше варианта, перевод численных значений в нечеткие производится без построения термов ЛП и расчета функций принадлежности). Попадание в определенный кластер позволяет отнести указанное значение к соответствующему терму.

На конечной итерации метода k-средних были получены уточненные значения центров масс кластеров.

Далее рассматривается отдельно каждый кластер (рис. 5). Рассчитывается нормированное расстояние точек кластера до его центра масс, вычтенное из единицы по формуле (4).

|  |  |
| --- | --- |
|  | (4) |

Полученные для каждого кластера отображаются в системе координат (T, *).*Т.е. эта операция позволила расположить значения температуры по степени близости к ядру кластера. Вычисленные расстояния представлены на рис. 6.

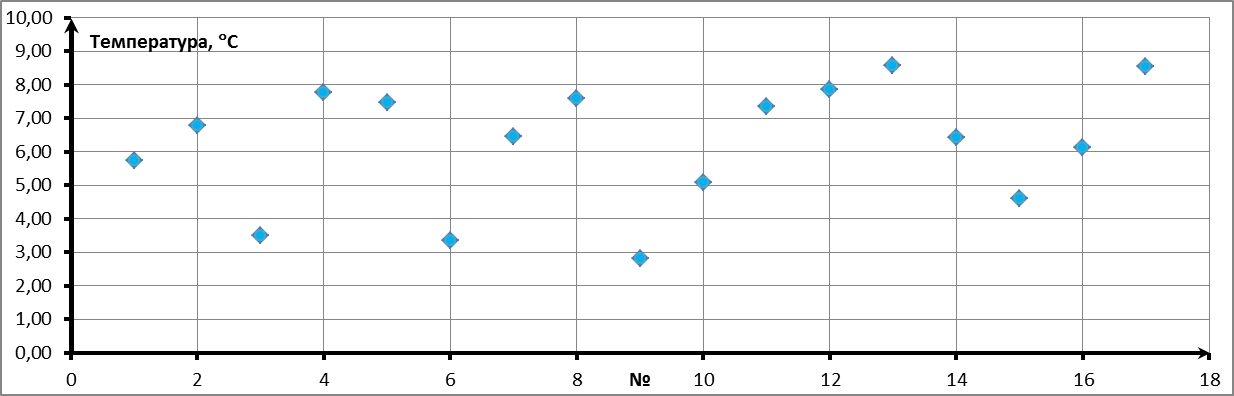
******

Рис. 5 – Кластер «Переходная»

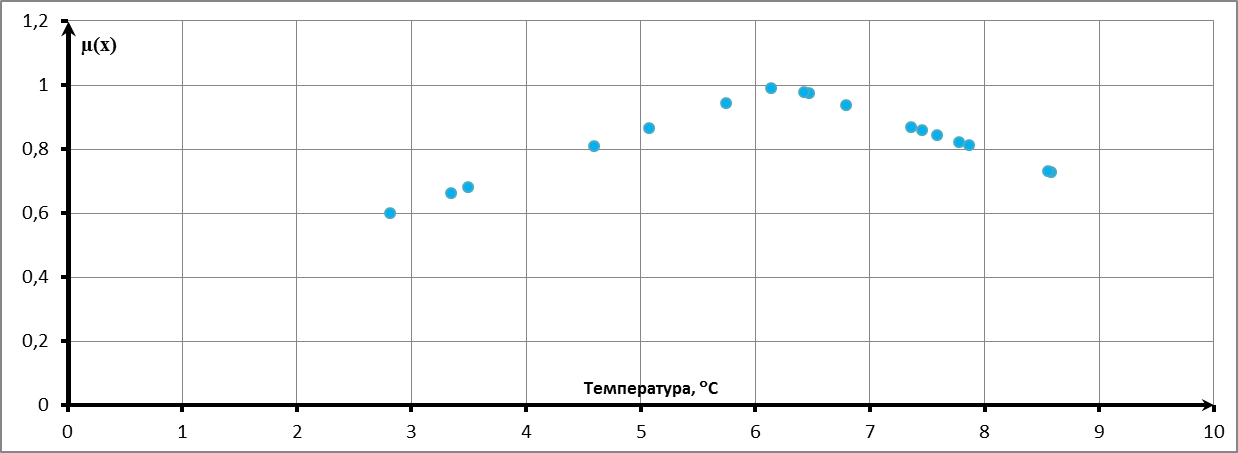
******

Рис. 6 – Кластер «Переходная» сортированный по близости точек к центру масс

Количество кластеров для множества соответствует количеству термов лингвистической переменной. Эти кластеры можно использовать для построения термов соответствующих лингвистических переменных. Предлагается провести аппроксимацию каждого кластера функцией вида:

|  |  |
| --- | --- |
|  | (4) |

Поскольку каждый кластер ставится в соответствие терму ЛП, полученную функцию можно представить в виде функции принадлежности соответствующего терма .

Функция вида (4) выбрана в соответствии с тем, чтобы не образовалось «пустых мест» в определении лингвистической переменной, в которых не будет определен ни один терм (так например будет, если аппроксимировать линейной функцией).

1. ***Автоматизация построения продукционных правил***

Для решения задачи автоматизации построения продукционных правил вида (1) предлагается решить задачу Data mining [5] – поиска ассоциативных связей.

*Постановка задачи поиска ассоциативных связей:*

Пусть *I = {i1, i2, i3, ...in}* - множество термов всех лингвистических переменных погодных условий. Пусть *J = {j1, j2, j3, ...jn}* множество термов лингвистической переменной «Количество лесных пожаров». Пусть *D* - множество строк таблицы, где каждая строка – это набор элементов из *I* и один элемент из *J*. Ассоциативным правилом называется импликация *X→Y*, где *X⊂I, Y⊂J*. Правило *X→Y* имеет поддержку *supp* (support), если доля строк из *D*, содержащая набор *X∪Y = supp*, *supp(X→Y) = supp(X∪Y)*.

Найти все правила *X→Y* с заданной минимальной поддержкой *supp*.

Для поиска ассоциативных связей воспользуемся алгоритмом Априори [5]. Алгоритм Априори предназначен для нахождения закономерностей (ассоциаций) между связанными событиями в базах данных.

*Описание алгоритма Априори:*

На первом этапе происходит формирование одноэлементных наборов (правил в инциденте которых только один элемент). Далее алгоритм подсчитывает поддержку (частоту) одноэлементных наборов. Поддержка рассчитывается по формуле (5)

|  |  |
| --- | --- |
| *,* | (5) |

где *S* – число строк, в которых встретился набор *X→Y;*

*D –* общее число строк.

Наборы с уровнем поддержки меньше установленного отсекаются. Оставшиеся наборы считаются часто встречающимися одноэлементными наборами. Далее происходит формирование двухэлементных наборов, подсчет их поддержки и отсечение наборов с заданным уровнем поддержки. и т.д. (Рис.7).

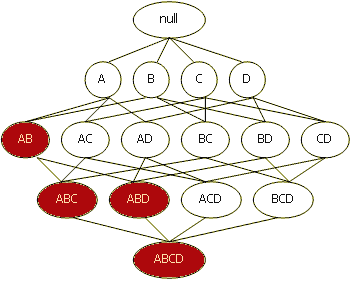


Рис.7 – Схема работы алгоритма Априори

Особенностью алгоритма Априори является то, что он уменьшает количество кандидатов, отсекая - априори - тех, которые заведомо не могут стать часто встречающимися, на основе информации об отсеченных кандидатах на предыдущих этапах работы алгоритма (свойство антимонотонности). То есть, если набор AC не является часто встречающимся набором, то и набор ACD не будет часто встречающимся, поэтому он не рассматривается в дальнейшем.

Исходная таблица 2 разбивается на n таблиц (в соответствии с числом термов лингвистической переменной «Количество лесных пожаров»). В каждой из этих таблиц реализовывается алгоритм Априори и следствием импликационных правил будет являться соответствующий терм лингвистической переменной «Количество лесных пожаров» таблицы (рис.8).



Исходная таблица данных

Рис.8 – Разбиение таблицы на подтаблицы

Т.е. в результате работы алгоритма для каждого терма ЛП «Количество лесных пожаров» формируется свой набор правил. Каждое правило формируется со своим весовым коэффициентом. В качестве такого коэффициента выступает значение поддержки (supp) правила.

Программная реализация алгоритма Априори для составления продукционных правил на языке C# [6] представлена на рис. 9. В программе есть возможность выбора уровней поиска (т.е. количества минимальных и максимальных составляющих правила) и регулировки поддержки для каждого уровня.

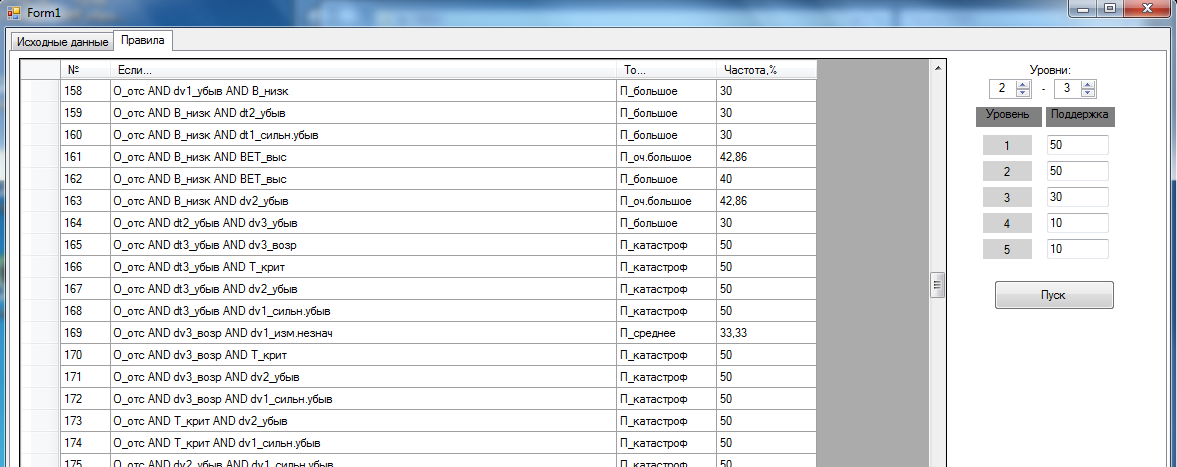


Рис.9 – Скриншот работы программной реализации алгоритма

***Вывод:***

Исходная таблица численных данных по метеоусловиям и по количеству лесных пожаров в районе при помощи кластеризации преобразовываются в нечеткую. Далее с помощью алгоритма Априори строятся нечеткие продукционные правила, которые ложатся в основу базы правил нечеткой продукционной системы. Затем формируются нечеткие лингвистические переменные и НПС дает прогноз количества лесных пожаров в районе на заданный интервал времени.

Автоматизация процесса составления функций принадлежности ЛП и поиска ассоциативных правил позволила наладить самостоятельную работу модели прогнозирования лесных пожаров, что исключит использование экспертов при ее построении.

Решение задачи прогнозирования количества лесных пожаров в районе позволит говорить о классе пожароопасности территории, путем дальнейшей обработки полученных данных. Классифицирование территории по степени пожароопасности, в свою очередь, позволит:

* принять превентивные меры к возникновению лесных пожаров (проливы водой, прокладка минерализованных полос, рубка и др.);
* принять меры к оповещению населения о возможной пожароопасной обстановке;
* заблоговременно определить состав сил и средств МЧС России для предупреждения и ликвидации ЧС.

**Литература:**

1. 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.
2. http: // rp5.ru / Архив погоды в Игнашино (дата обращения: 03.10.2012).
3. Леоненков А.В. Нечѐткое 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. 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.
6. Герберт Шилдт C# 4.0. Полное руководство. – Вильямс, 2013. – 1056 с.: ил.