**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 тыс. лесных пожаров, охватывающих площади до 2,5 млн. гектар. Ими государству ежегодно наносится ущерб в миллиарды рублей. Например, в 2013 году ущерб составил порядка 20 млрд.[1]

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. 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, сухость воздуха (дефицит влажности), скорости ветра, изменение этих параметров со временем и других[2].

В данной статье рассматривается третий этап математической модели прогнозирования количества лесных пожаров[3], а именно поиск зависимости пожаров от погодных условий и составление ассоциативных правил для оценки количества лесных пожаров.

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

**Исходные данные**

Таблица 1

Исходные данные

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Дата** | **T,°** | **ΔT2,°** | **ΔT3,°** | **V,%** | **ΔV2,%** | **ΔV3,%** | **…** | **F** |
| 01.04.12-08.04.12 | 5 | 3 | 5 | 65 | 10 | 18 | … | 22 |
| 09.04.12-  16.04.12 | 10 | 5 | 7 | 70 | 5 | 29 | … | 17 |
| … | … | … | … | … | … | … | … | … |
| 21.06.13-  28.06.13 | 15 | 8 | 9 | 80 | -6 | -13 | … | 0 |

Исходные данные задачи представляют собой набор временных рядов. В них входят усреднённые значения измерений в течение недели температуры, влажности, скорости ветра, других погодных условий и их модификации. Рассматриваются изменения температуры и влажности за три, за две и за одну неделю. Эти параметры введены в рассмотрение для того, чтобы учесть обусловлено тем, что долгое сохранение низкой влажности и высокой температуры ведёт к высыханию почвы и повышению возможности возгорания.

В соответствие этим данным ставится количество пожаров, зарегистрированных с помощью космического мониторинга (см. Табл.1).

**Постановка задачи**

Ассоциативные правила – это правила вида

Если А, то В (1)

Где А – это предпосылка, т.е. в нашем случае совокупность погодных условий данного региона, а В – результат, т.е. количество пожаров соответствующее условиям. Например, «Если температура высокая и влажность низкая, то количество пожаров будет высоким». В данной работе будет рассмотрен статистический метод построения продукционных правил.

Разобьём исходные данные на k подтаблиц, где k – мощность лингвистической переменной «Пожары» так, чтобы в каждой таблице были данные принадлежащие одному терму пожаров. Терму катастрофическое количество пожаров, очень высокое, высокое, среднее или низкое число пожаров. Т.е. погодные условия при катастрофическом числе пожаров попадают в одну таблицу, при высоком – в другую и так далее.

Далее для каждой таблицы решим задачу поиска ассоциативных правил.

Пусть *I = {i1, i2, i3, ...ik}* - множество термов всех лингвистических переменных погодных условий. Пусть *J = {j1, j2, j3, ...jk}* 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)*. Иными словами это частота встречаемости набора *X∪Y* в таблице.

Задача состоит в том, чтобы найти все правила *X→Y* с заданной минимальной поддержкой *supp*.

Найденные правила являются левой частью продукционного правила, частота встречаемости (поддержка) является достоверностью правила, а название терма пожаров для рассматриваемой таблицы – правой частью правила. В итоге получим правила вида (1).

**Алгоритм Apriori**

Для нахождения частых наборов используется алгоритм интеллектуального анализа данных Apriori. Главным его плюсом является использование свойства антимонотонности, позволяющее не учитывать заведомо редкие наборы. Т.е. наборы поддержка которых ниже заданного порога. Так же алгоритм отличается простотой реализации, что позволяет снизить издержки на его реализацию.

Проиллюстрируем работу алгоритма на абстрактном примере (Табл. 2). Пусть есть 4 переменных, каждая из которых принимает два значения 1 и 0. Найдём наиболее часто встречающиеся наборы единиц.

Таблица 2

Входные данные алгоритма

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| А | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Б | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| В | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| Г | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |

Первым этапом алгоритма Apriori является построение дерева частых наборов (Рис. 3).

Сначала создаётся пустая корневая вершина и в её потомки записываются все возможные термы базы данных (Рис. 3а). Здесь это А,Б,В,Г.

Дальше до тех пор, пока это возможно к каждой вершине добавляются все вершины, лежащие на том же уровне и правее. Т.е. к вершине А добавляются вершины Б, В, Г. К вершине Б соответственно В и Г, к вершине В – вершина Г (Рис. 3б).

В процессе этого для каждой вершины вычисляется значение её поддержки. Поддержка – это частота встречаемости набора в базе данных. Где набор состоит из самой вершины и её пути до корня. Например, терм «А» встречается в базе 3 из 10 раз, значит его поддержка равна 0.3. Набор «АБ» встречается один из 10 раз, значит поддержка равна 0.1. И если это значение меньше заданного порога (в данном случае он равен 0.15), то вершина отбрасывается (Рис. 3в). Так как все наборы, составленные из этой и любых других вершин, заведомо будут встречаться так же или менее часто в силу свойства *антимонотонности*. Которое гласит, что при добавлении к набору дополнительного элемента, его поддержка не может увеличиться. Например, при добавлении к набору АВ элемента Г, частота набора может либо не измениться, либо уменьшиться.



Рисунок 3. Первый этап алгоритма Apriori

Вторым этапом работы алгоритма является обход дерева кандидатов в глубину для составления наборов. Каждый путь от корня к листу дерева является наиболее часто встречающимся набором. Т.е. предпосылкой к правилу. Результатом же правила записывается количество пожаров, соответствующее текущей таблице.



Рисунок 4. Второй этап алгоритма Apriori

Таким образом основные составляющие модели прогнозирования пожарной обстановки построены. Количество пожаров вычисляется при помощи алгоритма прямого нечёткого вывода Мамдани.

**Апробация модели**

На данный момент для прогноза были выбраны следующие данные: средняя температура за неделю, средняя влажность за неделю и их прирост за одну, две и три недели, количество осадков в неделю и скорость ветра.

Описанная модель была реализована на языке программирования C++ в среде Qt Creator. Для апробации модели были выбраны данные о пожарах и погодных условиях по Сковородинскому району Амурской области за пять лет с 2008 по 2012. Были автоматически построены лингвистические переменные соответствующие погодным условиям региона и пожарам, а так же продукционные правила. Примеры построенных правил приведены в таблице 5.

Таблица 5

Примеры правил

|  |
| --- |
| Если , то количество пожаров *низкое* |
| Если , то количество пожаров *катастрофическое* |
| Если , то количество пожаров *низкое* |

На рисунке 4 представлены результаты автоматической аппроксимации пожароопасной обстановки в регионе.

*Рисунок 4. Аппроксимация данных по пожарам значениями полученными …*

Средняя погрешность прогноза при данных результатах составляет 21 пожар. А суммарная ошибка по всем 106 неделям 2295 пожаров. Такой прогноз нельзя назвать точным.

Было принято решение изменить параметры кластеризации и разбить пожары на четыре кластера. В первый поместить все недели, когда число пожаров было равно 0. А остальные недели равномерно разбить на три группы с низким, средним и высоким количеством пожаров. Так же в ходе исследования было выявлено, что наилучший прогноз даёт набор из 5 правил с наибольшей поддержкой. Увеличение правил ведёт к тому, что в нечётком выводе срабатывают правила с меньшей поддержкой и, следовательно, приводящие к ошибкам. Так же предметом исследования являлась задача выяснить, какого размера брать правила, т.е. какое количество термов должны быть в выбранных правилах. Наилучшее приближение дают правила с максимальным числом термов. Таким образом удалось снизить среднюю ошибку до 6 пожаров. А суммарную до 593. График прогноза представлен на рисунке 5.

*Рисунок 5. Прогноз после изменения параметров модели*

**Список литературы**

1. <http://ria.ru/eco_news/20131213/983924582.html>
2. «Прогнозирование количества лесных пожаров по данным космического мониторинга». Дрожжин, Белоусов.
3. Статья №3.