**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 прошлом году ущерб составил порядка 20 млрд.[1]

При этом 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 лесных пожаров в регионе.

Очевидно, что горимость лесов в большой степени зависит от температуры и сухости воздуха (дефицита влажности): чем выше температура и сухость воздуха, тем больше горимость и наоборот. Так же на пожароопасную обстановку влияет скорость ветра, количество осадков и изменение этих параметров со временем. Исходя из этого, целью данного исследования является поиск зависимости количества лесных пожаров от погодных условий и, соответственно, прогноз пожароопасной обстановки.

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

**Исходные данные**

Исходные данные задачи представляют собой набор временных рядов. В них входят усреднённые значения измерений в течение недели температуры, влажности, скорости ветра, других погодных условий наблюдаемой среды и их модификации. В соответствие им ставится количество пожаров, зарегистрированных с помощью космического мониторинга (см. Табл.1).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 |

*Таблица 1. Исходные данные*

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.

**Математическая постановка задачи**

Поэтому данная работа будет основана на методах нечёткой логики и нечёткого вывода. Алгоритм программы состоит из трёх этапов:

1. Построение лингвистических переменных, описывающих исходные данные и предобработка входных данных.
2. Поиск ассоциативных правил для оценки количества лесных пожаров.
3. Прямой нечёткий вывод, основанный на алгоритме Мамдани.

Первый этап состоит в том, чтобы построить функции принадлежности значений каждой переменной к терму лингвистической переменной. Обычно в теории нечётких множеств эта операция проводится вручную на основе субъективного представления эксперта о принадлежности конкретного значения тому или иному терму. В данной работе исследуется вопрос о возможности автоматического построения функций принадлежности при помощи кластеризации статистической информации о каждом из параметров среды.

Описание кластеризации?

Следующим этапом после приведения данных к качественному виду является поиск продукционных правил. Это правила вида «Если А, то В». 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.

**Поиск ассоциативных правил**

В первую очередь необходимо найти зависимость количества пожаров от погодных условий. Для этого разобьём данные на k подтаблиц, где k – мощность лингвистической переменной «Пожары» так, чтобы в каждой таблице были данные принадлежащие одному терму пожаров. Таблица, в которой записаны погодные условия при катастрофическом количестве пожаров, при очень высоком, высоком, среднем и низком.

В каждой таблице найдём все возможные часто встречающиеся наборы погодных условий. Совокупность этих условий является левой частью продукционного правила, частота встречаемости (поддержка) является достоверностью правила, а величина пожаров для рассматриваемой таблицы – правой частью правила.

Для нахождения частых наборов используется алгоритм интеллектуального анализа данных Apriori. Главным плюсом которого является использование свойства антимонотонности, позволяющее не учитывать заведомо редкие правила. Так же алгоритм отличается простотой реализации, что позволяет снизить издержки на его реализацию.

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

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

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

В первую очередь создаётся пустая корневая вершина и в её потомки записываются все возможные лингвистические значения 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.

*Таблица 2. Входные данные алгоритма*



*Рисунок 3. Алгоритм Apriori*

Последним этапом работы алгоритма является обход дерева кандидатов в глубину для составления правил. 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.

Таким образом основные составляющие модели прогнозирования пожарной обстановки построены. Количество пожаров вычисляется при помощи алгоритма прямого нечёткого вывода Мамдани. Перейдём к апробации модели на реальных данных.

**Апробация модели**

Описанная модель была реализована на языке программирования C++ в среде Qt Creator. Для данных по сковородинскому району Амурской области за четыре года с 2008 по 2012 были автоматически построены лингвистические переменные соответствующие погодным условиям региона и пожарам, а так же продукционные правила. Дальнейшая работа состояла в том, чтобы изменяя параметры кластеризации или проводя её вручную, а так же изменяя количество используемых правил добиться наилучшего приближения.

На рисунке 4 представлены результаты автоматического прогнозирования пожароопасной обстановки в регионе.

*Рисунок 4. Автоматический прогноз*

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

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*