ОБЪЯСНЯЙ! А не рассказывай. Как в школе.

**Слайд 1.**

Товарищ полковник, уважаемые коллеги, вашему вниманию предоставляется доклад рядового Костенчука Михаила на тему «Поиск ассоциативных правил для оценки количества лесных пожаров в модели ANFIS»

**Слайд 2.**

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.

При этом из анализа данных росстата, показанных на экране, следует, что количество пожаров на территории России за последние 20 лет не имеет тенденции к уменьшению.

**Слайд 3.**

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

Одним из этапов математической модели прогнозирования пожаров является использование ассоциативных правил, выявляющих зависимость пожаров от погодных условий. Поэтому целью данной работы является найти продукционные правила оценки количества лесных пожаров.

А задачей – реализация алгоритма Apriori.

**Слайд 4.**

Исходные данные задачи представляют собой набор временных рядов усреднённых значений 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.

**Слайд 5.**

Прежде чем перейти к решению поставленной задачи необходимо произвести предварительную обработку данных.

Для этого нужно перевести количественные характеристики в качественные. Сопоставив каждой характеристике среды одно из 5 лингвистических значений. Например, температура может быть очень\_низкой, низкой, средней, высокой и очень высокой. В дальнейшем для удобства мы сопоставляем каждому лингвистическому значению цифру от 1 до 5, где 1 – самое низкое значение, а 5 – самое высокое.

Об этом этапе подробно рассказывалось в предыдущем докладе.

**Слайд 6.**

Перейдём к решению задачи. В первую очередь необходимо понять, что такое ассоциативное правило. Это импликативное правило 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».

**Слайд 7.**

Нахождение правил осуществляется при помощи алгоритма интеллектуального анализа данных Apriori. Главными плюсами алгоритма являются

1. Свойство антимонотонности, позволяющее не учитывать заведомо редкие правила.
2. А так же простота реализации.

Работа алгоритма состоит из трёх этапов.

**Слайд 8.**

На первом этапе все данные разбиваются на 5 таблиц по количеству пожаров. Таблица, в которой записаны погодные условия при катастрофическом количестве пожаров, при очень высоком, высоком, среднем и низком.

Дальнейшая работа идёт с каждой таблицей отдельно.

**Слайд 9.**

Возьмём для примера одну из таких таблиц – с высоким числом пожаров. И для краткости сократим её, заменив погодные условия на абстрактные символы «А,Б,В,Г» и оставив в ячейках таблицы только два значения. Например, 1 – температура высокая, 0 – температура низкая. Основным этапом алгоритма Apriori является построение дерева частых наборов.

В первую очередь создаётся пустая корневая вершина и в её потомки записываются все возможные лингвистические значения базы данных. Здесь это А,Б,В,Г.

**Слайд 10.**

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 с термами Б, В, Г. А к вершине Б соответственно В и Г. К вершине В – вершина Г.

**Слайд 11.**

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 15%), то вершина отбрасывается, т.к. 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.

**Слайд 12.**

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.

**Слайд 13.**

В результате апробации алгоритма на исходных данных получены следующие правила, которые совпадают с интуитивном представлении о зависимости количества пожаров от погодных условий: …

Найденные правила в дальнейшем используются в математической модели прогнозирования пожаров.

На этом мой доклад окончен, спасибо за внимание.