ОБЪЯСНЯЙ! А не рассказывай. Как в воскресной школе.

**Слайд 1.**

Здравствуйте, уважаемые товарищи. 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 от 0,5 до 2,5 млн. га. Ими государству ежегодно наносится ущерб в миллиарды рублей. Например, в 2013 году ущерб составил порядка 20 млрд рублей.

При этом из анализа данных федеральной службы государственной статистики, 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 построение ассоциативных правил, выявляющих зависимость пожаров от погодных условий. Поэтому целью своей работы я ставлю найти связи и построить продукционные правила оценки количества лесных пожаров.

Для достижения этой цели необходимо выполнить следующие задачи:

1. Предобработка входных данных.
2. Реализация алгоритма Apriori.

**Слайд 4.**

Исходные данные представляют собой набор временных рядов усреднённых значений измерений в течение недели температуры, влажности, скорости ветра и других погодных условий наблюдаемой среды и их модификации, в соответствие им ставится количество пожаров, зарегистрированных с помощью космического мониторинга.

**Слайд 5.** Перехожу к изложению методов решения поставленных задач.

Первая задача состоит в проведении предобработки исходных данных, т.е. необходимо перевести количественные характеристики в качественные. Например количественной характеристике температуры в 5° соответствует качественная «средняя температура». А температуре в -3 градуса значение «очень низкая температура». Об этом этапе подробно рассказывалось в предыдущем докладе.

Хотелось бы отдельно заметить, что на данном этапе разработки модели мощность характеристик погодных условий выбрана равной пяти. Т.е. каждому показателю присваивается пять термов по шкале от 1 до 5, где 1 – это самое низкое значение, а 5 – самое высокое. Например, температура может быть очень низкой, низкой, средней, высокой и очень высокой.

**Слайд 6.**

Второй задачей является нахождение ассоциативных правил. Т.е. импликативных правил, 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» или «Если температура низкая и влажность низкая, то количество пожаров будет средним».

**Слайд 7.**

Нахождение правил осуществляется при помощи алгоритма интеллектуального анализа данных Apriori. Алгоритм удобен тем, что

1. Использование свойства антимонотонности наборов, которое позволяет не учитывать заведомо редкие правила, существенно уменьшает время работы программы в сравнении с простым перебором.
2. А так же простота реализации.

Работа алгоритма состоит из трёх этапов.

**Слайд 8.**

На первом этапе все данные разбиваются на 5 таблиц с одинаковыми правыми частями правила. В нашем случае правой частью (результатом правила) является количество пожаров. Таблица, в которой записаны условия при которых количество пожаров катастрофическое, таблица, в которой количество пожаров высокое и так далее.

Дальнейшая работа идёт с каждой таблицей отдельно.

**Слайд 9.**

Следующим и основным этапом является построение дерева частых наборов.

В первую очередь создаётся пустая корневая вершина и в её потомки записываются все термы базы данных. Например, температура высокая, температура низкая, влажность высокая и так далее. На слайде они обозначены условными буквами А, Б, В, Г.

**Слайд 10.**

Дальше до тех пор пока это возможно к каждой вершине добавляются все термы лежащие на том же уровне и правее, чем эта вершина. 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.**

Следующим этапом для каждой вершины вычисляется значение её поддержки. Т.е. 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. Это свойство называется *антимонотонностью*. При добавлении к набору дополнительного элемента, его поддержка не может увеличиться.

**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: …

Полученные правила используются в модели прогнозирования пожаров.

**Слайд 14.**

Таким образом искомая цель достигнута и работа выполнена.