

ИСПОЛЬЗОВАНИЕ МЕТОДА АГЛОМЕРАТИВНОЙ ИЕРАРХИЧЕСКОЙ КЛАССИФИКАЦИИ ПРИ РАЗДЕЛЕНИИ МИКРОСИТУАЦИЙ НА КЛАССЫ ЛАВИНООПАСНЫХ И НЕЛАВИНООПАСНЫХ

Е.В. Яковлева

Харьковский национальный университет радиоэлектроники

e-mail: vitamin_K@ukr.net

Казалось бы, колов, присуций снегу, должен был сообщить ему оцепенелость зимы, а белизна – неподвижность савана. Однако это опровергается стремительным движением лавины. Лавина – это снег, ставший огненной печью. Она ледяная, но все пожирает.

Виктор Гюго, «Человек, который смеется»

Лавиноопасные районы занимают на Земле около 6% площади суши — 9253 тыс.кв.км. И одной из важнейших задач при освоении лавиноопасных районов являются противолавинные мероприятия: защита от отдельных лавинных очагов и предупреждение попадания в лавины людей. Если просчитать экономическую обоснованность противолавинных мероприятий, то становится ясно, что прогнозирование лавин во много раз дешевле затрат на уборку снега, применение снеговыводящих заборов, защитных щитов, противолавинных сеток, посадку защитных лесонасаждений. Обрушение лавин происходит при нарушении устойчивости снежного покрова на склоне, вызванном воздействием внешних факторов и процессами внутри снежной толщи.

Каждая микроситуация, характеризующая метеообстановку в лавиноопасном районе, характеризуется набором факторов (предикторов). При классификации этих факторов принято выделять лавинообразующие – те, которые влияют на сход лавины, - и нелавинообразующие – те, которые на сход лавины не влияют. В зависимости от конкретного района эти факторы могут быть различными. Также следует отметить, что на процесс лавинообразования оказывают влияние не только отдельные факторы, но и их сочетания. В связи с этим, факторы делят на зависимые и независимые.

В данной работе ставится задача разделения групп метеоданных – микроситуаций – на два класса: лавиноопасные и нелавиноопасные.

Так как значения предикторов могут быть измерены в разных шкалах (как количественных, так и качественных), то целесообразно нормализовать матрицу, например, по методу Саати, на основе суждений экспертов.

Допустим, уже сформирована матрица $A_{n \times m}$ с элементами $a_{ij} > 0, i = 1 \dots n, j = 1 \dots m$, где m – число микроситуаций, а n – число предикторов, характеризующих микроситуацию.

На первом шаге считаем, что каждый объект является отдельным кластером.

Начальное действие агломеративной процедуры состоит в вычислении всех мер

сходства d_{ij} . Используя метрику с весами $d(x, y) = \sqrt{\sum_{i=1}^n \alpha_i (x_i - y_i)^2}$, где α_i - весовой

коэффициент, характеризующий важность (вес предиктора), $\alpha_i \in [0;1]$, (устанавливается экспертами), а $\bar{x}(x_1, x_2, \dots, x_n)$ и $\bar{y}(y_1, y_2, \dots, y_n)$ - микроситуации для сравнения, составим матрицу расстояний $G_{n \times n}$ между всеми микроситуациями:

$$G = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}, \text{ где } d_{ij} = d_{ji}, i = 1 \dots n, j = 1 \dots m$$

Пара объектов с наибольшей мерой сходства, т.е. $\min[d_{ij}]$, объединяется в новый $(n+1)$ -ый кластер. Затем определяется подходящая мера сходства между этим новым кластером и остальными $n-2$ микроситуациями. На каждом шаге размерность матрицы расстояний G сокращается на единицу и осуществляется то объединение (между двумя микроситуациями, между микроситуацией и кластером или между двумя кластерами) для которых мера сходства максимальна, т.е. $d_{ij} = \min_{i,j} d_{ij}, i \neq j$. На p -ом шаге объединений повторяется та же процедура на матрице $G_{(n-p) \times (n-p)}$ пока все микроситуации не объединятся в один кластер. Всего в процессе работы процедуры вычисляется $(n-1)^2$ мер сходства.

Если в процессе работы процедуры на каком-либо шаге сразу между несколькими микроситуациями (кластерами) имеются одинаковые минимальные расстояния, то алгоритм разветвляется на два пути:

- 1) выбирают одну случайную близкую пару микроситуаций (классический подход) или
- 2) объединяют в одну группу все близкие микроситуации.

Результаты работы оформляются в виде дендрограммы (рис. 1), по горизонтали которой показаны номера микроситуаций, а по вертикали значения межкластерных расстояний d_{ij} , при которых произошло объединение двух данных кластеров.

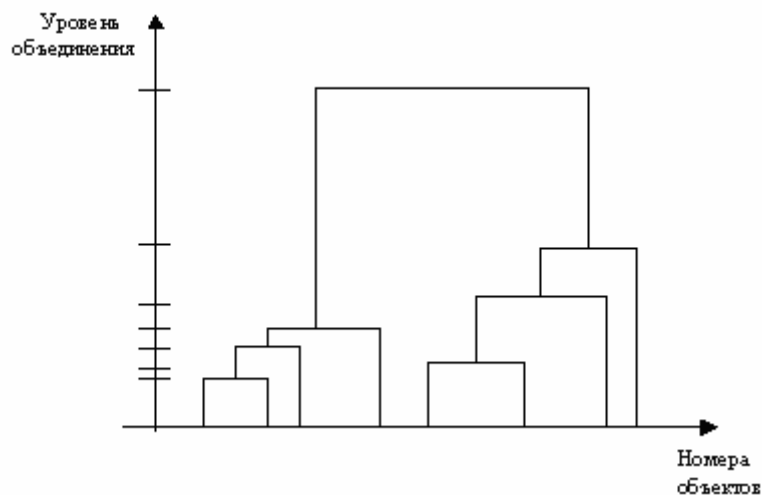


Рис. 1

На некотором шаге объединения похожих микроситуаций в кластер возникает ситуация, когда необходимо объединить кластер с микроситуацией или кластер с кластером.

Пусть имеется два кластера i и j с количеством элементов n_i и n_j , соответственно. Допустим, что d_{ij} - минимальное расстояние из всех оставшихся, так что i и j объединяются и образуют новый кластер $h = (i \cup j)$ с $n_i + n_j = n_h$ числом элементов. Тогда расстояние между кластерами h и l (h получен объединением i и j в единый кластер):

$$d_{l(i \cup j)} = d_{lh} = \alpha_1 d_{il} + \alpha_2 d_{jl} + \beta d_{ij} + \gamma |d_{il} - d_{jl}|, \quad (1)$$

где $\alpha_1, \alpha_2, \beta, \gamma$ - константы, значения которых, определяют вид стратегии объединения.

Воспользуемся монотонной стратегией средней невзвешенной связи, при которой коэффициенты равны $\alpha_1 = \alpha_2 = 0.5, \beta = \gamma = 0$.

Тогда расстояние между кластерами h и l примет вид:

$$d_{l(i \cup j)} = d_{lh} = 0.5d_{il} + 0.5d_{jl}$$

На некотором шаге все микроситуации и кластеры объединятся в один кластер с очень малой мерой сходства. Устанавливая на основании суждений экспертов некоторое пороговое значение, меньшее по значению, чем мера сходства для данного кластера, получим искомые классы лавиноопасных и нелавиноопасных микроситуаций. Конечно, существует вероятность отнесения микроситуации не к тому классу, существуют ошибки первого и второго рода, но также есть возможность минимизации вероятности этих ошибок с последующим уточнением принадлежности микроситуации к какому-либо классу.

Литература

1. Мандель М.Д. Кластерный анализ, М.: Финансы и статистика, 1988.
2. Факторный, дискриминантный и кластерный анализ / Под ред. Енюкова И.С., М.: Финансы и статистика, 1989.
3. Дюран Б., Оделла П. Кластерный анализ.- М.: Статистика, 1977.