

## ОБ ОДНОМ КРИТЕРИИ ДЛЯ ВЫБОРА ЗНАЧЕНИЯ ЭКСПОНЕНЦИАЛЬНОГО ВЕСА В АЛГОРИТМЕ КЛАССИФИКАЦИИ НЕЧЕТКИХ С-СРЕДНИХ

Смешко Ю. В.,

научный руководитель канд. техн. наук Агафонов Е. Д.

*Сибирский Государственный Аэрокосмический университет имени академика*

*М. Ф. Решетнева*

*Сибирский Федеральный университет*

Процесс классификации является первоочередным этапом интеллектуального анализа данных. В задаче классификации с самообучением исходная информация об объектах представлена признаковым описанием без указаний учителя. Количество классов заранее неизвестно и определяется в процессе группировки. При нечетком подходе к классификации класс представляет собой нечеткое множество объектов, границы которого размыты. Наиболее распространенная постановка задачи классификации с самообучением в терминах нечетких множеств сводится к оптимизации функционала, предложенного Дж. Беждеком – Дж. Данном:

$$Q(P) = \sum_{l=1}^c \sum_{i=1}^m \mu_{li}^{\gamma} d(x_i, \tau^l) \rightarrow \min, \quad (1)$$

где  $\mu_{li}$  - функция принадлежности  $i$ -объекта  $l$ -классу;  $d(x_i, \tau^l)$  - функция, определяющее расстояние от  $i$ -объекта до центра  $l$ -класса;  $\gamma$  - экспоненциальный вес;  $c$  - количество классов ( $l = 1, 2, \dots, c$ );  $m$  - количество объектов ( $i = 1, 2, \dots, m$ ). Смысл оптимизации функционала (1) заключается в минимизации нечетких отклонений объектов от центров нечетких классов.

Алгоритм оптимизации критерия (1) называется алгоритмом нечетких  $S$ -средних. Настраиваемым параметром алгоритма является экспоненциальный вес  $\gamma$ , для которого выполняется условие  $1 \leq \gamma \leq \infty$ . Параметр  $\gamma$  отвечает за степень размытия выделяемых в процессе классификации классов. При  $\gamma = 1$  полученное разбиение является четким, то есть каждый объект исследуемой совокупности оказывается однозначно принадлежащим какому-либо одному классу. С увеличением значения  $\gamma$  возрастает неопределенность классификации, которая выражается следующим образом:

$$\gamma \rightarrow \infty \Rightarrow \mu_{li} \rightarrow \frac{1}{c} \quad \forall l = 1, 2, \dots, c; \forall i = 1, 2, \dots, m. \quad (2)$$

Известно несколько способов настройки параметра  $\gamma$ , которые предлагают определенным образом ограничивать интервал варьирования значения экспоненциального веса. Выбор значения параметра внутри найденного интервала является актуальной задачей.

В работе предлагается критерий, с помощью которого можно количественно оценить степень размытия полученного разбиения. Исходная информация для расчета критерия задана матрицей  $F_{m \times c} = \{\mu_{li}\}$ , которая определяется в результате группировки  $m$  объектов в  $c$  нечетких классов с помощью алгоритма нечетких  $S$ -средних.

Предлагается следующий алгоритм для расчета критерия:

1. Для каждого объекта исходной совокупности определяем максимальную функцию принадлежности и назначаем объекту метку, соответствующую порядковому номеру нечеткого класса с которым данный объект имеет максимальную степень сходства:

$$\forall i = 1, 2, \dots, m: \text{Label}_i = \{l : \max_{1 \leq l \leq c} (\mu_{li})\}. \quad (3)$$

2. Для каждого класса рассчитываем количество появлений его порядкового номера среди элементов вектора  $\text{Label}_i$ , (то есть, определяем элементы вектора  $\text{Kolobjects}_i$ , которые показывают, сколько объектов отнесено к каждому классу):

$$\forall l = 1, 2, \dots, c: \text{Если } \text{Label}_i = l, \text{ Тогда } \text{Kolobjects}_l = \text{Kolobjects}_l + 1. \quad (4)$$

3. Для каждого класса рассчитываем его нечеткий вес:

$$\forall l = 1, 2, \dots, c: \text{Vesclassa}_l = \sum_{i=1}^m \mu_{li}. \quad (5)$$

Предлагается следующий критерий для количественной оценки степени размытия полученного разбиения:

$$W = \frac{1}{c} \sum_{l=1}^c |\text{Kolobjects}_l - \text{Vesclassa}_l|. \quad (6)$$

В качестве исходных данных для алгоритма использовалась выборка Glass Identification, взятая с UCI Machine Learning Repository. Шесть типов стекла представлены 214 образцами. Для каждого образца измерено 9 различных показателей-признаков. Признаки выражены количественными величинами. Указания учителя отсутствуют. Исходное множество объектов разделено на 6 классов алгоритмом нечетких  $C$ -средних. Различные варианты нечеткой классификации в многомерном признаковом пространстве представлены линейными диаграммами.

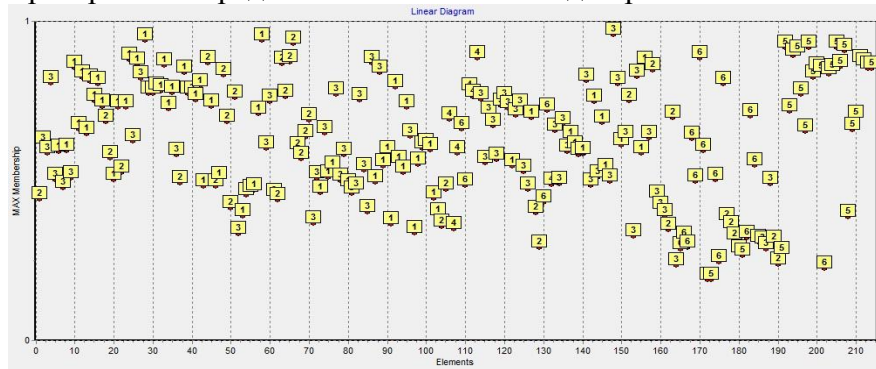


Рис.1. Размытое разбиение ( $\gamma = 2, W = 1.713$ )

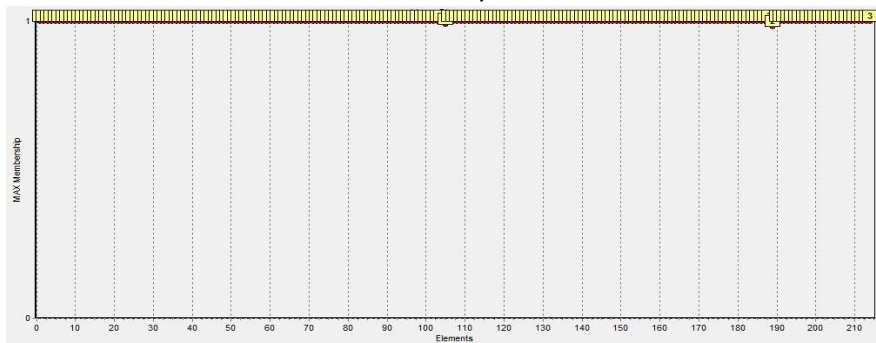


Рис.2. Однозначное разбиение ( $\gamma = 1.02, W = 0.003$ )

Значение критерия (6) зависит от выбора экспоненциального веса  $\gamma$ .  $W < 1$  указывает на практически однозначное разбиение, то есть каждый объект исходного множества оказался отнесен только к одному классу.  $W > 1$  приводит к размытой классификации, при этом величина критерия (6) является количественной оценкой степени размытия построенного нечеткого разбиения.