

Гучук Владимир Всеволодович,
к.т.н., ст.н.с. ИПУ РАН, Москва

ТЕХНОЛОГИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ ПРОЦЕДУРЫ ПОВЫШЕНИЯ КАЧЕСТВА КЛАСТЕРИЗАЦИИ ПО ЭКСПЕРТНЫМ ОЦЕНКАМ

Аннотация: Описываются особенности реализации интерактивной процедуры объективизации, разработанной автором и позволяющей на основе анализа измеряемых параметров улучшать качество кластеризации многопараметрических объектов по экспертным оценкам. Процедура использует свойство полиаттракции – неравномерное распределение значений векторов, описывающих реальные объекты в пространстве измеряемых параметров. При этом параметры как бы группируются вокруг нескольких координатных значений, возможно классообразующих. Для осуществления процедуры объективизации необходимо определиться с количеством измеряемых параметров объектов, после чего на основе оценок привлекаемых экспертов с учетом их квалификации и специализации определенная часть объектов будет отнесена к одному из классов, количество которых определяется в процессе выполнения этого этапа на основе полученных данных. Затем осуществляется ранжирование векторов по степени принадлежности к своему классу, присвоенному объекту экспертными оценками. Далее осуществляется объективизация, заключающаяся в применении к экспертным оценкам процедуры корректировки результатов оценивания. Приводится блок-схема алгоритма.

Ключевые слова: многопараметрические объекты, экспертные оценки, полиаттракция, класс, кластеризация, объективизация, формализация.

В статье анализируются технологические особенности реализации интерактивной процедуры, предложенной автором в [1], которая направлена на повышение качества кластеризации сложных объектов по экспертным оценкам [2, 3]. Описывается последовательность обработки имеющихся данных для проведения процедуры – процедуры объективизации, заключающейся в применении к экспертным оценкам процедуры корректировки результатов оценивания. Приводится блок-схема алгоритма корректировки.

Описываемая процедура объективизации использует свойство полиаттракции – неравномерное распределение значений векторов $[\vec{V}_1 \dots \vec{V}_N]$, описывающих реальные объекты $[O_1 \dots O_N]$ в пространстве измеряемых параметров $[P_1 \dots P_L]$. Как следствие – параметры как бы группируются вокруг нескольких координатных значений $\vec{X}_i$, $i = 1 \dots Q$, возможно классообразующих. При этом достаточно большое число объектов достаточно уверенно оцениваются при экспертном оценивании как относящиеся к одному из типов T_i , $i = 1 \dots M$, объектов. В общем случае $M \neq Q$.

Если достаточно часто совпадает экспертное отнесение объекта к конкретному типу объекта T_j и попадание его параметров в компактную область конкретного координатного значения $\vec{X}_j$, то выполняется известная гипотеза компактности [4].

Для проведения вышеописанных процедур необходимо определиться с измеряемыми параметрами $[P_1 \dots P_R]$ объектов, после чего на основе оценок экспертов с учетом их квалификации и специализации определенная часть объектов будет отнесена к одному из M классов. В процессе выполнения этого этапа должна определиться величина M , а также



осуществиться общий анализ полученных данных. После этого следует произвести ранжирование каждого из N векторов по степени принадлежности к своему i -му классу, который был присвоен объекту экспертными оценками. На первом подэтапе из массива векторов каждого класса случайным образом выделяется некоторая часть векторов, составляющих обучающую выборку (ОВ). Относительно сложностей формирования ОВ см [5, 6]. Далее, на основе ОВ находится L наиболее эффективных параметров для осуществления идентификации. Затем идентифицируются остальные вектора, попавших в тестовую выборку (ТВ). Далее последовательность описанных процедур, начиная с формирования ОВ, циклически повторяется. Число циклов зависит от объема выборки. На каждом цикле, для каждого вектора $\vec{V}_n$ подсчитывается текущее (суммарное) число S_n попаданий в ТВ и, если идентификационная оценка E класса вектора совпадает с экспертной оценкой $E_{\vec{V}_n}$ класса, увеличивается текущее (суммарное) число $S_{n,i}$ правильных идентификаций: $\forall n (\vec{V}_n \in \text{ТВ}), (S_n = S_n + 1) \& (E\vec{V}_n = E_{\vec{V}_n} \rightarrow S_{n,i} = S_{n,i} + 1)$. В итоге вычисляется кластерный коэффициент принадлежности к своему i -му классу $K_{n,i} = S_{n,i} / S_n$. Вычисляются также степени принадлежности к чужим классам $K_{n,j \neq i}''$. Второй подэтап аналогичен первому. Отличие заключается в том, что ОВ формируется уже из векторов, имеющих высокие предварительные кластерные коэффициенты. Вычисляются кластерные коэффициенты, после чего вектора ранжируются согласно этим кластерным коэффициентам. Такая методология напоминает идею бустинга, описанную в [7]. Далее осуществляется непосредственно объективизация, заключающаяся в применении к экспертным оценкам процедуры корректировки результатов оценивания, в том числе переклассификация, основанная на системе:

$$\left\{ \begin{array}{l} (\exists i, (\vec{V}_n \in \{\vec{V}\}_i) (K_{n,i} < K_*)) \rightarrow ((\exists j, K_{n,j} > K^{**}) (\forall j, K_{n,j \neq i} < K_{**})) \rightarrow (\vec{V}_n \in \{\vec{V}\}_j); \\ (\forall i, (\vec{V}_n \notin \{\vec{V}\}_i) \rightarrow ((\exists j, K_{n,j} > K^{**}) (\forall l, K_{n,l \neq j} < K_{**})) \rightarrow (\vec{V}_n \in \{\vec{V}\}_j); \\ (\exists i, (\vec{V}_n \in \{\vec{V}\}_i) (K_{n,i} < K_*)) \rightarrow ((\exists j, K_{n,j} > K^{**}) (\forall l, K_{n,l \neq j} > K_{**})) \rightarrow (\forall m, \vec{V}_n \notin \{\vec{V}\}_m); \\ (\exists i, (\vec{V}_n \in \{\vec{V}\}_i) (K_* \leq K_{n,i} \leq K^*)) \rightarrow (\exists j, K_{n,j} > K^{**}) \rightarrow (\forall m, \vec{V}_n \notin \{\vec{V}\}_m). \end{array} \right.$$

Здесь: N – число векторов, n – порядковый номер вектора $\vec{V}_n$, $\{\vec{V}\}_i$ – множество векторов i -го класса (класс вектора определяется согласно экспертной оценке или предыдущей переклассификации), $K_{n,i}$ – коэффициент принадлежности n -го вектора к i -му классу, K^{**} , K^* , K^{**} , K^* – минимальные и максимальные пороговые значения для коэффициентов принадлежности. Область значений ниже K^* (K^{**}) соответствует отсутствию выраженных признаков конкретного класса, а область значений выше K^* (K^{**}) – явному наличию признаков класса. Значения с одним “звездным” индексом используются при анализе степени принадлежности к своему классу, в противном случае используется другое пороговое значение. В результате объективизации отсеиваются ошибочные решения экспертов, повышается достоверность результатов кластеризации и формируются более адекватные формализованные методы идентификации экспериментальных данных. Упрощенная блок-схема алгоритма переклассификации представлена на рис. 1.

Определяющее влияние на эффективность процедуры объективизации, оказывают несколько факторов: – факторы внутреннего или формального характера (представительность выборки векторов и валидность измеряемых параметров [8]). – факторы внешнего (привносимого) плана (потенциальная возможность экспертного оценивания и качество экспертных оценок [9], зависящее от квалификации экспертов).



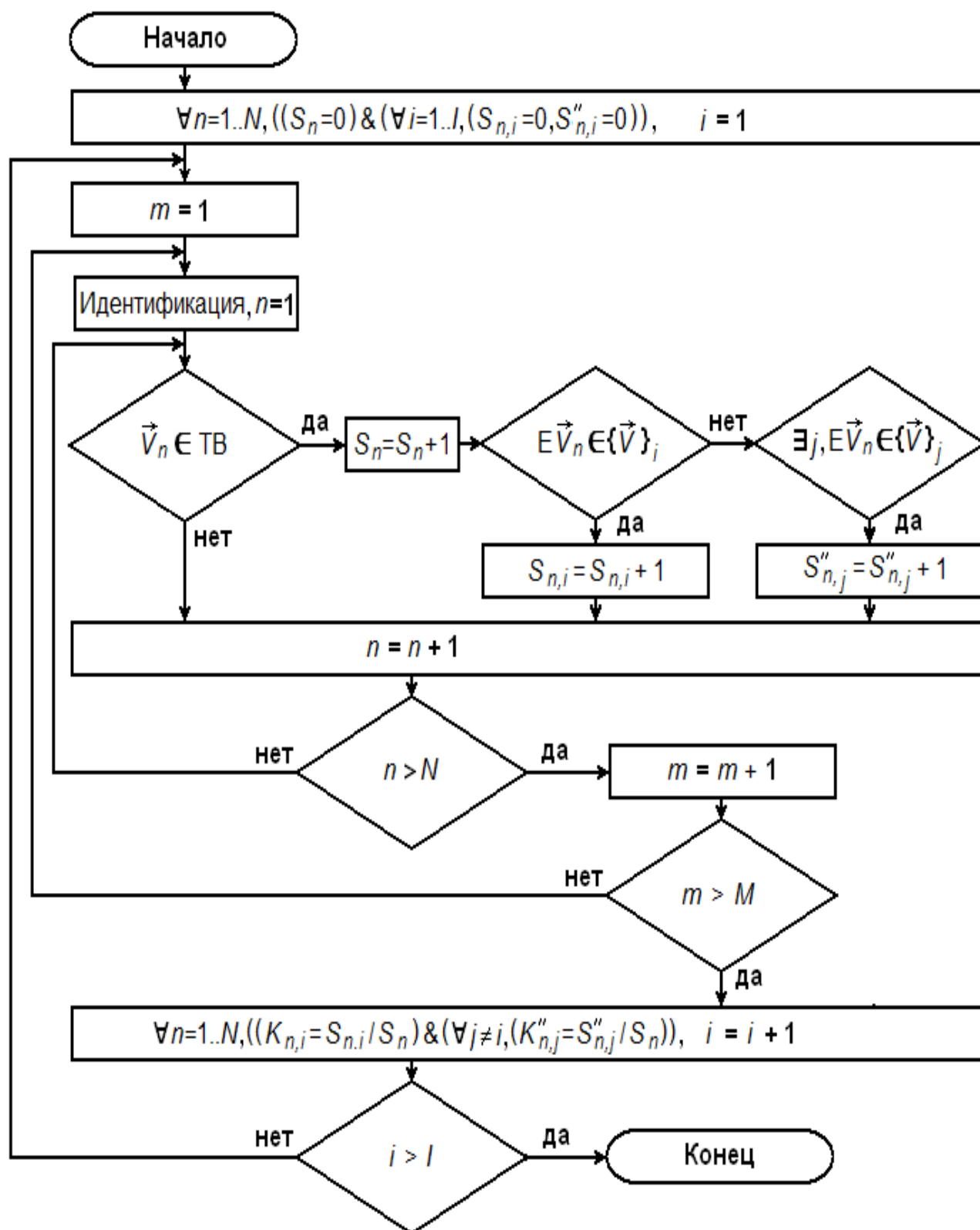


Рис. 1. Упрощенная блок-схема алгоритма автоматической переклассификации.

Работы по объективизации экспертных оценок проводились и ранее [10, 11]. Процедуры разрабатывались для уточнения выставленных в ранговых шкалах экспертных оценок качества одной группы объектов и их можно использовать внутри процедуры объективизации экспертной кластеризации.

Рассматриваемая процедура объективизации может быть использована, например, при создании алгоритмов медицинской диагностики по пульсовым сигналам лучевой артерии на основе субъективной кластеризации формы пульсовых сигналов [12].



Список литературы:

1. Guchuk V.V. Application of algorithms of objectifying expert clustering of Multiparameter objects in the analysis of big arrays of information // *Advances in Systems Science and Applications*. – 2018. – 18 (1). – с. 102-109.
2. Мандель И. Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
3. Гуцыкова С. В. Метод экспертных оценок. Теория и практика. – М: Когито-Центр, 2011. – 144 с.
4. Аркадьев А. Г., Браверман Э. М. Обучение машины классификации объектов. – М.: Физматгиз, 1971. – 192 с.
5. Кафтанников И. Л., Парасич А. В. Проблемы формирования обучающей выборки в задачах машинного обучения // *Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника*. – 2016. – 16 (3). – с. 15-24.
6. Ефимова И. В. Формирование однородных обучающих выборок для задач медицинской диагностики // *Труды 57-ой международной научной конференции МФТИ*. – 2014. – с. 91-92.
7. Friedman J. Greedy function approximation: a gradient boosting machine // *Annals of Statistics*. – 2001. – 29 (5). – p. 1189-1232.
8. Количественное описание неопределённости в аналитических измерениях / Под ред. Р.Л. Кадиса. – СПб.: ЦОП «Профессия», 2016. – 158 с.
9. Литвак Б. Г. Экспертные оценки и принятие решений. – М.: Патент, 1996. – 271 с.
10. Strijov V., Krymova E., Weber G. W. Evidence optimization for consequently generated models // *Mathematical and Computer Modelling*. – 2013. – 57 (1–2). – p. 50-56.
11. Гушин А. В. Алгоритм объективизации оценок эксперта с помощью измеряемых данных // 51-я науч. конф. МФТИ: конф.: Тр. конф. М.: МФТИ. – 2009. – с. 138.
12. Гучук В.В., Покровская И.В., Дорофеюк А.А., Десова А.А. Интеллектуальный анализ квазипериодических биосигналов в задачах медицинской диагностики (на примере пульсового сигнала) // *Автоматика и телемеханика*. – 2018. – 11. – с. 3-15.

