Об изучении структуры организационной системы с помощью метода вычисления оценок

Мгеладзе А. П.
Грузинский Технический Университет, г, Тбилиси

Резюме

В работе описываются методы вычисления оценок, с целью построения кластер-анализа для эмпирической информации об организационных системах с помощью выше указанного метода.

Ключевые слова: *Метод, вычисления оценок, множество признаков, число голосов, организационная система, близость, сходства.*

В настоящее время одним из общепринятых подходов к изучению оргсистем является подход, основанный на использовании тех или иных процедур кластеризации. Вместе с тем, имеется силный разрыв между огромными массивами эмпирической информации о конкретных оргсистемах и возможностями обработки этих массивов указанными процедурами кластеризации.

Одна из коренных причин этого разрыва состоит в том, что имеющаяся эмпирическая информация является разнородной, включающая наряду с числовыми, номинальными и ранговыми признаками, также признаки более сложной структуры. В то же время, наиболее апробированные процедуры кластерного анализа – это процедуры обработки информации чисто числового характера. В редком числе изветны процедуры кластеризации, ориентированныена номинальные или ранговые данные. Причем из этих последних процедур одни работают только с номинальными данными, а другие – только с ранговыми.

В связи со сказанным представляет большую актуальность изучить методы кластеризации общего типа, не зависящие от типа используемых признаков и могущих обрабатывать разнотипные данные, характерные для описания оргсистем. Среди таких методов наиболее приспособленными, по нашему мнению, являются методы вычисления оценок. Не уступая методам ориентированным на анализ данных одного типа в вычислительном отношении, методы вычисления оценок превосходят последние по легкости интерпретации результатов, возможностям участь разную априорную информацию, резерву адаптивности к новым требованиям и задачам.

Цель данного реферата состоит в том, чтобы описать эти методы (их основную схему) с целью обосновать целесообразновть их применения для анализа оргсистем.

Метод вычисления оценок – один из наиболее изученных методам решения задачи распознавания образов и таксоновии. Число исследований, посвященных его развитию, составляет много сотен. Однако выразительные возможности составляющих его элементов далеко не исчерпаны. Их наглядность и согласованность с интуитивными представлениями прикладников различных профилей делает этот метод уникальным

полигоном для изобретателей новых задач анализа структуры эмпирических данных и новых алгоритмов решения этих задач.

Реферат преследует цель придания черт адаптивности алгоритмам, реализующим этот метод, и, тем самым, расширения сферы применения метода на области с малой априорной информацией о структуре изучаемых классификаций и информационной базы, на которую опираются процедуры определения принадлежности лбъектов к классам этих классификаций.

После монографии Ю. И. Журавлева, суммировавшего основные алгоритмические достижения метода, конструктивное развитие этого метода замедлилось. До недавного времени казалось, что его развитие исчерпало свои возможности. Предлагаемый в реферате подход показывает, что это представление не соответствует действительности.

В последние годы ведется большая работа по упорядочению разных методов и алгоритмов классификации. В связи с этой работой неоднократно высказывалась точка зрения, что число предложенных методов и алгоритмов классификации, в которой прикладники могут только запутаться.

Не умаляя важности упорядочивающих исследований, мы не разделяем опасений, ожидаемых от чрезмерного богатства методов и алгоритмов классификации. Мы считаем, что этих методов и алгоритмов мало; мало для того, чтобы охватить хотя бы основные из бытующих в практике содержательные представления о сходстве и различии, о классах и классификациях. Поэтому работа по «изобретению» новых методов и алгоритмов остается актуальной. Этот поисх Модельной характеризации одного из важнейших творческих актов изучения действительности человеком. В этом смысле работа по созданию новых методов и алгоритмов классификации – важное направление в области искусственного интеллекта.

Алгоритмы распознавания, основанные на методе вычисления оценок, реализуются обычно как оптимизационные процедуры. В последние годы в связи с интенсивной разработкой теоретического (алгебраического) подхода к синтезу алгоритмов основные усилия были направлены на создание наиболее общих процедур такого рода. Между тем специфика «наивной» формы представления метода вычисления оценок позволяет конструировать совершенно новые алгоритмы, которые хотя формально можно рассматривать как частные реализации общих процедур, отличаются большой содержательной наглядностью, имеющей в приложениях не меньшее значение, чем количественные критерии эффективности.

Этой специфике метода вычисления оценок до сих пор уделялось недостаточное внимание. В самом деле, самые характерные элементы методи вычисления оценок (и, одновременно, наиболее содержательные с интуитивной точки зрения элементы) – это функция r_ω, определяющая бинарное отношение безразличия между объектами на подмножестве признаков ω, это так называемая система $\Omega(\omega \in \Omega)$ опорных подмножеств признаков, и, наконец, это функция $\Gamma(K,S)$– число голосов, которое отдает класс К за объект S (степень принадлежности S к К). И во всех предлоденных до сих пор алгоритмах эти наиболее характерные элементы по сущетву не оптимизировались. Именно, при оптимизации системы Ω варьировалась лишь мощность ее элементов, выбираемая одинаковой для всех элементов, а при оптимизации функции $\Gamma(K,S)$ варьировался лишь вектор весо в голосующих объектов из К.

Такое положение не случайно, т. К. До недавнего времени оптимизация $r_{\omega,\Omega}$ и Г даже не выделялись как самостоятельные задачи, не осознавались как таковые. К тому же отсутсвовала формальная постановка этих задач. Естественно поэтому, что попытка их решения были очень несовершенны.

В данном реферате подробно разбирается только одна из указанных задач, - задач побдора оптимальной в определенном смысле системы Ω, - и предлагается ряд алгоритмов ее решения[1]). В рамках наивного описания метода вычисления оценок выбор конкретной системы Ω означает выбор конкретного варианта алгоритма, реализующего этот метод. Варьируя этой системой, можно добиться требуемого изменения алгоритма классификации. Поэтому описываемая далее версия метода вычисления оценок, основанная на активном подборе системы Ω, была названа адаптивной. В отличие от своего прообраза (обычной версии метода вычисления оценок), в котором оптимизация системы: Ω велась или неявно (варьированием алфавита признаков) или явно, но в очень ограниченных пределах (изменением мощности элементов) в адаптивный метод расходует основные вычислительные затраты на прямой активный анализ широкого допустимого семейства систем Ω.

Указание возможности конструктивного задания разных вариантов достаточно широких семейств систем Ω как областей поиска, удобных для организации оптимизационных процедур, - главная особенность предлагаемой версии метода вычисления оценок. Рассматриваются два типа такого задания – параметрический и непараметрический. В первом случае точки допустимого семейства отображаются в точки специального пространства параметров, и процедура оптимизации реализуется в сконструированном пространстве параметров; во втором случае преобрезованию подвергается непосредственно система Ω (одни ее элементы заменяются на другие).

При обычном подходе к построению алгоритма вычисления оценок стремился к такому выбору системы Ω,при котором вычисление значений функции Г можно было проводить в свернутой форме, не суммиря явно «голоса», полученные на элементах этой системы. Отсюда и возникали ограничения на параметризацию допустимых систем Ω, не позволяющие хорошо анализировать структуру исходного пространства признаков. По этой же причини непараметрические вариации системы Ω вовсе не рассматривались.

В предлагаемой версии метода, наоборот, проблема упрощения вычисленний значений функции Г сознательно оставляется без внимания. Предполагается, что на любой допустимой системе эти значения можно определять непосредственно. Это приводит, конечно, к сильному ограничению на мошность /Ω/ рассматриваемых систем. Зато достигается большая свобода в организации процедуры варьирования такими системами. Более того, сильное ограничение на мошность /Ω/ систем выступает не как стеснительное условие, а как удобный фильтр, позволяющий в ходе решения задачи выявить ценность исходной информации. Кроме того, системы с малой мошностью оказываются простыми в интерпретации получаемых результатов. Учитывая, что в последнее время задачи обучения распознованию и таксономии решаются большой

[1] Это не совсем точно. В работе рассматриваются некоторые способы варьирования функциями r_{ω} и Г. На основе этого рассмотрения предложены даже некоторые алгоритмические конструкции целенаправленного изменения этих функций. Однако все это выполнено «по ходу дела», без целенаправленного и детального исследования возможностей этих вариаций для придания алгоритмам вычисления оценок свойства адаптивности.

частью не столько ради нахождения решающих правил классификации, сколько ради оценки возможностей тех или иных групп признаков к дифференциации классов, следует рассматривать простоту интерпретитуемости результатов как важную характеристику формируемых алгоритмов.

Эта характеристика особенно важна в задачах таксономии, где оценка качества (критерий) очень грубо отражает цели исследователя. В данной работе задаче таксономии уделено значительное место, что предопределило особое внимание к тому, чтобы допустимые системы Ω строились с учетом требования легкой интерпретируемости.

Неоднакратно отмечалось, что варьирование системой Ω (при фиксированных функциях r_ω и Г) в методе вычисления оценок эквивалентно варьированию метрикой в метрических методах классификации. С этой точки зрения предлагаемый адаптивный подход имеет в качетве предшественника метод раздвигающей метрики, предложенный себастьянов и развитый В.П. Якубовичем и Ю. И. Неймарком. Наиболее близка к нему схема построения таксономии с согласованной метрикой.

Одна из неявных целей реферата состоит в том, чтобы возродить интерес к «изобратению» новых алгоритмов вычисления оценок, орипающихся на наглядные элементы наивной формы представления метода. Это позволит, мы надеемся, объединить в общих процедурах интуитивную ясность и малое число степеней свободы эвристических алгоритмов и универсальность и обоснованность теоретических конструкций.

Как и многие другие методы агрегирования данных метод вычисления оценок опирается на введение функции близности между объектами. Однако, в отличие от общепринятого подхода, когда такая функция или заимствуется из теории метрических пространств или конструируется для специального случая, в методе вычисления оценок это введение организовано как некоторая общая схема анализа структуры пространства исходных признаков. Необходимость конкретизации элементов этой схемы при введении функции близости требует от исследовантеля активной оценки всей априорной информации относительно структуры исходного пространства, которой он располагает. Особенноть конструкции этих элементов допускает учет свойств априорной информации самого различного типа. Таким образом, метод вычисления оценок-это метод, специально ориентированный на синтез эмпирической информации о конкретных объектах и общей информации о классах таких объектов.

Центральным элементом схемы введения функции близности в методе вычисления оценок является функция сходства частей объектов. Из множества P исходных признаков выделим некоторое подмножество $\omega(\omega \subseteq P)$. Совокупность значений координат объекта S, соответствующих подмножеству ω, называется ω – частью этого объекта. На множестве M_ω ω – частей всех рассматриваемых объектов определяется бинарное отношение $R_\omega(S,S')$ неразличимости, удовлетворяющее условиям рефлексивности и симметричности. Искомая функция $r_\omega(S,S')$ сходства между ω – частями объектов S и S' задается в виде

$$r_\omega(S,S')=\begin{cases} I, \text{ если между } \omega \text{ частями } S \text{ и } S' \\ \text{имеется место отношение } R_\omega(S,S'), \\ 0, \text{ в противном случае} \end{cases} \quad (1)$$

Таким образом, функция сходства – это характериктическая функция отношения неразличимости, определенная на квадрате M_ω^2 множества M_ω.

Пусть теперь тем или иным способом в P выделено некоторое семейство Ω подмножеств признаков и на каждом элементе ω этого семейства $\omega \in \Omega$ определено (вобще говоря, свое) отношение R_ω(S,S′). Каждая пара объектов S и S′ разбивает это семейство на два непересекающихся подсемейства Ω^0(S,S′) и Ω′(S,S′). :

а) $\omega \in \Omega^0$(S,S′), если r_ω(S,S′)=0

б) $\omega \in \Omega'$(S,S′), если r_ω(S,S′)=1

Очевидно, что $\Omega=\Omega^0 \cup \Omega'$. В методе вычисления оценок в качестве функции f(S,S′) близости между двумя объектами предлагается выбирать мошность /Ω′(S,S′)/ множества Ω′(S,S′). Она, оевидна, равна числу характеристических функций соответствующих отношений Rω (S,S′) неразличимости, которые включают пару (S,S′) . это число можно записать в виде суммы

$$f(S,S')=\sum_{\omega\in\Omega} r_\omega(S,S') \tag{2}$$

Семейство Ω, с помощью которого вычисляется функция (2), называется системой опорных подмножеств.

С помощью функции (2) функция Г(K,S), оценивающая степень принадлежности объекта S к произвольному конечному множеству K объектом, того же пространства (клссу K) определяется просто как сумма:

$$\Gamma(K,S)=\sum_{S'\in K} f(S,S') \tag{3}$$

Часто вместо (3) используется нормированная величина:

$$\Gamma^n(K,S)=\frac{1}{|K|}\bullet\Gamma(K,S), \tag{4}$$

где модуль /K/ как обычно обозначает мошность множества K.

Функцию (3) (или 4)) называют мерой близости между объектом S и множестов K; в работах по методу вычисления оценок она иногда называется функцией голосования (говорят, что она определяет число голосов, которые отдают объекты S′ из класса K за объект S).

Следует подчеркнуть, что бинарное отношение Rω(S,S′), которое задается вместе с каждым элементом ω (и которое как раз определяет значения функции r_ω(S,S′)) может быть конкретизовано очень большим числом способов. Более того, мы можем отказаться от условий рефлексивности и симметричности и выбиреть Rω из всего множества всех бинарных отношений. Лишь бы выбранное отношение адекватным образом моделировало требуемый смысл противопоставления представлений «сходство-различие» сравниваемых объектов на подпространстве ω (можно рассматривать и более широкую область выбора для Rω – например, класс всех бинарных отношений с весами).

Естественно, чтобы выбор Rω (задание области выбора) не может не быть зависим от природы признаков, входящих в рассматриваемое опорное подмножество ω, но зависимость тоже может быть разной.

Большие возможности вырьирования имеются и при выборе системы Ω. возможны вариация и в задании функции f и Г. Эту последнюю возможность проиллюстрируем простыми примерами:

$$f(S,S') = \sum_{\omega \in \Omega} \alpha_\omega \bullet r_\omega(S,S'), \tag{5}$$

$$\Gamma(K,S) = \sum_{S' \in K} \beta_{S'} \bullet f(S,S'), \tag{6}$$

где введены вектора «весов» $\{\alpha_1,.....,\alpha_{|\Omega|}\}, u\{\beta_{1,}.....,\beta_{|K|}\}$ Для характеризации разной «важности» между элементами из Ω и объектами из К соответственно.

Возможности и варьирования элементов метода вычисления оценок, на которые было указано, с одной стороны, делают этот метод очень гибким, но, с другой стороны,эти возможности слишком разнообразны, чтобы исследователь «на глазок» до обработки правильно фиксировал эти элементы. Именно поэтому уже в первых работах по созданию метода алгоритмы конструировались как оптимизационные процедуры, которые автоматически подбирали требуемые элементы в заданной области варьирования. Однако задание таких областей было несовершенно, и, как уже отмесалось, настоящее исследование стремится преодолеть, прежде всего, именно это несовершенство.

Рассмотрим случай, когда все признаки являются двоичными. В этом случае часто используются два простых варианта выбора системы Ω опорных подмножеств:

а) множество Ωn всех одиночных признаков (Ωn=P: $|\Omega_n| = |P| = n)$, ,

б) множество Ω всех возможных подмножеств множества P. В обоих вариантах полагается, что

$$r_\omega(S,S') = \begin{cases} 1, \text{ если S совпадает с S′ на } \omega \\ 0, \text{ противном случае} \end{cases} \tag{7}$$

Тогда для первого варианта получаем

$$f_1(S,S') = n - \rho(S,S') \tag{8}$$

а для второго

$$f_2(S,S') = 2^{n-\rho(S,S')} - 1, \tag{9}$$

где через $\rho(S,S')$ обозначено расстояние Хемминга между векторами с двоичными координатами (число несовпадающих разрядов у сравниваемых векторов). Функция f_1 – линейная функция, а f_2 – существенно нелинейная функция расстояния ρ. Построение функции Г(К,S) на базе f_1 означает равноправное сравнение больших и малых расстояний. В частности, пропорциональное увеличение расстояний между всеми объектами из М (т.е. пропорциональное растяжение структуры объектов в заданном пространстве Р) не влияет на отношение разных значений этой функции. Оно сказывается только на ее абсолютных значениях. Напротив, использование функции f_2 означает, что делается акцент на учет расстояний, которые не превышают некоторого заранее выбранного эффективного уровня.

Между функциями f_1 и f_2 имеется целый ряд промежуточных по сложности функций. В качестве системы Ω опорных подмножеств выбираются все подмножества мошности не больше некоторого заданного числа К(К<n):

$$\Omega_{np} = \bigcup_{i=1}^{K} \Omega_i, \tag{10}$$

где через Ωi обозначно семейство всех подмножеств признаков мощности i, (то есть, если ω∈Ωi, то /ω/=i), а функция rω(S,S′) сходства для каждого ω∈Ωnp определяется в соответствии с (7). Построенная на базе такого семейства функция близности $f_{np}(S,S')$ имеет вид:

$$f_{np}(S,S') = \sum_{i=1`}^{\tilde{\tilde{K}}} C^{i}{}_{n-\rho(S,S')}, \qquad (11)$$

где через $C^{i}{}_{n-\rho(S,S')}$ обозначено число сочетаний из $n-\rho(S,S')$ по i, а число $\tilde{K}$ определяется из условия

$$\tilde{K} = \begin{cases} \text{K, если } K \le n-\rho \\ \text{n-ρ, если K> n-ρ} \end{cases} \qquad (12)$$

Как это видно из (11), при K=1 имеем $f_{np} = f_1$, а при K=n функция $f_{np} = f_2$. При малых K функция f_{np} ведет себя пропорционально $(n-\rho)^K$, то есть аналогично полиному K-й степени от ρ, значения которого убывают по мере приближения ρ к n со скоростьою порядка ρ^K.

Вычисление функций f_{np} не требует явного перебора выбранной системы Ω опорных множеств и поэтому не является сложным. Более того, всегда имеется возможность перейти от формул (11), требующих подсчета факториалов, к другим более простым формулам, которые, приближенно сохраняя смысл соотношений (11), не требуют подсчета факторялов. Например, возможно использовать просто полиномы от ρ заданной степени. Вместе с тем, явный учет происхождения f_{np} как функции, составленной на базе выбранной сист. Ω опорных подмножества имеет свои преимущества. В частности, может представить интерес исключение из данного семейства Ωi подмножеств признаков мошности i некоторой его небольшой части Ωi′, мошность которой так мала, что вычислительно нетрудно просмотреть эту часть явно. Например, легко предположить, что в случае K=2 желательнл сравнивать объекты по такой функции близости, которая строится на базе всех подмножеств мошности «единица», но не всех подмножеств мошности «два». Те подмножества мошности «два», которые требуется исключить, могут быть указаны выделением тех 2-3 признаков, сочетание которых с другими признаками, по мнению специалистов, не является информативным.

В разобранных случаях предполагалось, что функция сходства $r_\omega(S,S')$ для всех ω∈Ω вычисляется одинаково по формуле (7), то есть на основе сравнения ω частей пары объектов на совпадение. Очевидные обобщения этой формулы связаны со смягчением условия неразличимости ω – частей сравниваемых объектов:

$$r'_\omega(S,S') = \begin{cases} 1, \text{ если } \rho_\omega(S,S') < \varepsilon_\omega \\ 0, \text{ в противном лучае} \end{cases} \qquad (13)$$

где ρ_ω – расстояние по Хеммингу ω – частей объектов S и S', а ε_ω – пороговая константа неразличимости объектов для данного множества ω признаков. Она выбирается заранее и является свободным параметром алгоритма ($\varepsilon_\omega \ge 0$).

Рассмотрим вариант использования формулы (13), в котором пороговая константа одинакова во всех подмножествах выбранного семейства Ω и равна ε. В этом случае, если выбрать Ω в виде $\Omega = \bigcup_{i=1}^{K} \Omega_i$, то получим следующее соотношение для определения

$$f_{np} = C^{n-\rho\omega(S,S')}_{n-\rho\omega(S,S')+\varepsilon} \bullet \sum_{i=1}^{\tilde{K}} C^{i}_{n-\rho},$$

где использованы те же обозначения, как и в (11), определяющей $f_{np}(S,S')$.

В монографии Ю. И. Журавлева по методу вычисления оценок подчеркивается, что имеется два принципиально разных способа введения системы Ω опорных множеств.

Первый способ состоит в том, что фиксируется заранее независимо от анализируемых эмпирических данных определеное свойство А выделяемых множеств. Если подмножество ω обладает этим свойством, то это модмножество включается в выделяеиую систему. Чтобы подчеркнуть, что системы выделяются с помощью заданного заранее мостоянного независящего от обрабатываемых данных свойства А, оно обозначается через Ω_A. Для того, чтобы задать свойство такого типа, обычно достаточно знать только число признаков в таблице T_{nme} обрабатываемых объектов, (n-число классов, на которые делятся объекты, причем для каждого объекта из T_{nme} известно, какому из этих классов он принадлежит).

Второй способ основан на использовании такого свойства, наличие которого у данного подмножества ω существенно зависит от конкретного наполнения таблицы T_{nme} и расматриваемой классификации К строк этой таблицы (описываемых ею объектов). Поэтому второй способ позволяет строить такие системы Ω опорных множеств, которые специально приспособлены для различения объектов на разные классы имеено данной классификации и именно данной таблицы T_{nme}. В качестве примеров реализации этого способа можно отметить выдаление систем так называемых тестов и тупиковых тестов.

Подмножество ω признаков относится к семейству $\Omega(T_{nme}, \mathscr{K})$ тестов классификации $\mathscr{K}$заданной на таблице T_{nme}, если для любых двух различных классов K_q и $Kp(q \neq p)$ из $\mathscr{K}(K_q, K_p \in K$)выполняется условие

$$r_\omega(S,S') = 0 \text{ где } S \in K_q, S' \in K_p, \tag{15}$$

где r_ω вычисляется в соответствии с (7).

Пусть М – множество рассматриваемых объектов (строк таблицы T_{nme}, /М/=м). Введем на множестве всех подмножеств множества Р признаков характеристическую функцию δ(ω, М) теста:

$$\delta(\omega, M) = \begin{cases} 1, \text{ если } \omega\text{-тест} \\ 0, \text{ в противном лучае} \end{cases} \tag{16}$$

где символ М подчеркивается, что определение этой характеристической функции зависит не только от ω, но и от множества М объектов, на котором изучаются признаки.

Очевидно, что если δ(ω,М)=1, то

1) $\delta(\omega, M/S)=1$
2) $\delta(\omega', M)=1$ для всех ω' таких, что $\omega \le \omega' \le P$

Первое из этих свойств теста говорит о том, что сужение исходного множества M объектов сохраняет все тесты, найденные на охватывающем множестве. Из второго свойства следует, что наибольший интерес представляют маломощные тесты. Из него также следует целасообразность выделенипя специального подкласса тестов, названных тупиковыми. Характеристическое свойство тупикового теста заключается в том, что никакое его собственное подсмножество не является тестом, то есть, если ω – тупиковый тест, то для всех $\omega' \subset \omega$, $\delta(\omega', M)=0$. Выделение тупикового теста означает выделение такого подпространства $\omega \in P$, которое является безизбыточным по отношению к заданным классификации $\mathcal{K}$ и таблице T_{nme}.

С точки зрения качества анализа выбор системы Ω опорных подмножеств в виде множества всех тестов или, еще лучше, множества тупиковых тестов является высокоэффективным. Однако, с точки зрения оценки сложности выполнения такой выбор является затруднительным или даже невозможным во многих практически интересных случаях (когда /P/~20÷40,/M/~100÷300). Построение такого рода систем требует осуществления слишком большого объема вычислений. Более того, алгоритм реализации такого выбора имеет принципиально переборный характер (экспоненциальную сложность). Особенно трудоемкими оказываются алгоритмы определения семейства тупиковых тестов. Практический интерес могут представить процедуры выделения не всех тестов, а только тестов малой мошности. При этом, целесообразно строить системы Ω так, чтобы в него не включались тривиальные тесты, которые охватывают уже найденные подмножества – тесты еще меньшей мошности (смотри (17) пункт «2»).

Главный недостаток процедур выделения тестов малой мощности состоит в том, что получающаяся системи Ω сама может оказаться маломощной, а иногда и вовсе пустой. Чтобы как-то ограничить этот недостаток прибегают к статистическим алгоритмам поиска тестов. В этом случае разрешается поиск тестов разной мошности, но число поисковых проб ограничено. Такие алгоритмы, не снижая объема вычислений на проверку свойства «тестовасти» отдельного подмножества признаков, резко упрощают организацию их перебора. Чтобы добиться снижения объема вычислений на проверку свойства «тестовости» отдельного подмножества используют следующие три модификации понятия теста.

Подмножество ω называется (q, p)- тестом, если в классификаций $\mathcal{K}$ найдется такая пара различных классов Kq и Kp, что на части таблицы T_{nme}, выделяемой объектами этих классов, ω является тестом в определенном выше обычном смысле. В этой модификации вместо тесттов в качестве элементов системы опорных подмножеств предлагается искать (q,p)-тесты.

Подмножество ω называется (α,β) – квазитестом, если, во-первых, вместо проверки условия (15) для всех S из M эта проверка осуществляется на заранее выбранной доле α мощности /M/ этого множества, причем выбыраемая $\alpha\bullet$ /M/ часть объектов распределяется по M не случайно, а в каждом из ℓ классов заданной классификации $\mathcal{K}$ отбирается число, равное α – части мощности этих классов; во-вторых, вместо того, чтобы для каждого выбранного S (пусть $S \in Kq$) осуществлять проверку (15) для всех $S' \in M \setminus Kq$, проводится проверка лишь на числе $\beta\bullet \left| M \setminus Kq \right|$ таких объектов.

Подмножество ω называется (γ,q) – представительным тестом для класса Kq, если, во – первых, ω –это тест в обычном (указанном выше) смысле, и, во – вторых, ω- часть не менее, чем у $\gamma \bullet |Kq|$ числа объектов класса Kq совпадает.

Указанные модификации поняния теста хотя и дают практическое снижение объема вычислений при проверке данного подмножества (является ли оно тестом), все же не делают это снижение гарантированным. Кроме того, получающееся снижение не оказывается очень большим (сокращает объем вычислений менее, чем в два раза). Наконец, это снижение достигается обычно снижением информационной ценности выделяемых подмножеств: условия для выделения (q,p) и (α, β) тестов – это смягчение условий (15).

Уже этих обстоятельств достаточно, чтобы заключить, что целесообразно создавать новые процедуры поиска контекстнозависимых систем опорных подмножеств, не связанных с понятием теста. Накопление определенного «запаса» таких процедур, представляет конечно, и самостоятельный интерес: каждая такая процедура строится на выявлении особой связи между классификацией, индуцированной заданной эмпирической матрицей T_{nme}, и структурой исходного множества признаков, в которой эта классификация функционирует. В диссертации предполагается построить много существенно различных примеров таких процедур, выявить соответствующие связи, и определить конкретно как эти процедуры должны работать для анализа данных об оргсистемах.

ЗАКЛЮЧЕНИЕ

Из сказанного ясно, что гибкость метода вычисления оценок очень высокая. Поэтому расплывчатость и неопределенность, присущая данным об оргсистемах, может быть наиболее просто компенсирована, если строить кластерный анализ наличной косвенной информации об этих системах именно методами вычисления оценок. Такая работа еще не проводилась. Откуда следует, что высказанный тезис нуждается в практической проверке. Более того, специфика данных об оргсистемах может потребовать существенного развития (приспособления) метода вычисления оценок, но, как видно из реферата, именно данный метод располагает соответствующим резервом.

Цитированная литература

Журавлёв Ю. И. Распознавание Классификация Прогноз. Москва «Наука» 1989

Статья получена: 2008-10-16