2. Смирнов А. Н. Основные тенденции развития рынка огнеупорных материалов [Электронный ресурс] URL: http://steellab.com.ua/news/2014/01/01.php.

3. Обзор рынка магнезиального сырья (магнезита и брусита) и магнезитовых порошков в СНГ (3-е изд.) / INFOMINE Research Group. М. : ИнфоМайн, 2011. 137 с.

4. Зырянова В. Н. Использование магнийсодержащих отходов в производстве строительных материалов: дис. …канд. техн. наук / В. Н. Зырянова. Новосибирск, 1987. 249 с.

5. Перепелицын В. А. Техногенное минеральное сырье Урала / В. А. Перепелицын, В. М. Рытвин, В. А. Коротеев и др. Екатеринбург : РИО УрО РАН, 2013. 332 с.

6. Хуснутдинов В. А. Физико-химические основы технологии переработки нетрадиционного магнезиального сырья на чистый оксид и другие соединений магния: дис. … д-ра техн. наук / В. А. Хуснутдинов. Казань, 2000. 434 с.

7. Прокофьева В. В. Магнезиальные силикаты в производстве строительной керамики / В. В. Прокофьева, З. В. Багаущинов. СПб. : Золотой орел, 2005. 160 с.

8. Хорошавин Л. Б. Магнезиальные огнеупоры: справочник / Л. Б. Хорошавин, В. А. Перепелицын, В. А. Кононов. М. : Интермет Инжиниринг, 2001. 576 с.

9. Kramer D. A. Current mining of olivione and serpentine // U. S. Geological Survey Open-Pile Report, Reston. Virginia, 2002. 256 p.

10. Shand M. A. The Chemistry and Technology of Magnesia. John Wiley&Sons, Inc., 2006. 263 p.

АЛГОРИТМЫ И СОЗДАНИЕ ПРОГРАММНОГО СРЕДСТВА ФОРМИРОВАНИЕ СИСТЕМЫ ОПОРНЫХ МНОЖЕСТВ ПРИЗНАКОВ, ОБЕСПЕЧИВАЮЩИЙ НАДЕЖНОСТЬ РАСПОЗНАВАНИЯ

Бекмуратов Косим Аллабердиевич
кандидат технических наук, доцент Самаркандского филиала Ташкентского университета информационных технологий,Самарканд,Узбекистан.
Холматов Орзумурод Абжалолович
магистрант Самаркандского филиала Ташкентского университета информационных технологий,Самарканд,Узбекистан.

ALGORITHMS AND SOFTWARE TOOLS FORMATION OF REFERENCE SET OF FEATURES PROVIDING RELIABLE RECOGNITION

Bekmuratov Kosim Allaberdiyevich,
associate professor of technical science Samarkand branch of Tashkent University of information technologies, Tashkent, Uzbekistan.
Xolmatov Orzumurod Abjalolovich,
master Samarkand branch of Tashkent University of information technologies, Tashkent, Uzbekistan..

АННОТАЦИЯ

В данной статье приводится методика нахождения предельнодопустимой размерности пространства комбинации признаков. Предложены принципы формирования системы опорных множеств с учетом предельнодопустимой размерности комбинации признаков относительно конкретного образа. Разработан алгоритм и программное обеспечение, обеспечивающее требуемые качества и надежность распознавания объектов.

ABSTRACT

In this article, the technique of finding the maximumallowable dimensions of the feature space. The principles of formation of the system of support sets based extremely allowable dimension combinations of features on a specific image. The algorithm, which provides the required object recognition for quality and objects.

Ключевые слова: объект, признак, класс, эталонная выборка, контрольная выборка, объем эталонной выборки, системы опорных множеств, качество и надежность распознавания, вероятность ошибки, предельная размерность пространства, сумма голосов, решающее правило.

Key words: object, sign, class, reference sample, the control sample, the volume of the reference sample, the system of support sets, quality and reliability recognition, the probability of error, the limiting dimension of the space, the amount of votes, the decision rule.

Постановка задачи. Существуют различные методы установления критериев, позволяющие определить информативность отделных признаков или их комбинаций. Однозначно отдать предпочтение одному из них невозможно. Методы, лучше работающие для одного из классов задач, хуже приемлемы для другого класса, вместе с тем, во многих методах при установлении критерия информативности признаков или их комбинаций не учитывается объем эталонной выборки (число признаков и объектов). Более того, при установлении критериев информативности признаков или их комбинаций необходимо учитывать такие важные параметры, как качество и надежность распознавания.

В [1-2] получены теоретические результаты, указывающие весьма осторожно относиться к большинству известных методов распознавания, не уделяющих специального внимания формированию признакового пространства, в котором в процессе обучения строятся решающие правила. Эти выводы можно рассматривать как обоснование главной цели настоящей статьи.

В [3] получен теоретический результат, смысл которого состоит в том, что если на эталонной выборке из N решающих правил выбирается одно, которое безошибочно разделяет эталонную выборку длины m, то с вероятностью $(1-\eta)$ можно утверждать, что вероятность ошибки при распознавании объектов с помощью этого правила составит величину, меньшую ε, где

$$\varepsilon = \frac{\ln N - \ln \eta}{m} . \qquad (1)$$

Из (1) следует, что чем проще решающее правило и чем ниже размерность признакого пространства, тем меньше вероятность ошибочных ответов, возникающих при эксплуатации распознающей системы.

В работах [4, 5] показано, что системы опорных множеств Ω_A строятся следующим образом:

a) $\Omega_A = \{\Omega : |\Omega| = k\} = C_n^k$;

b) $\Omega_A = \{\Omega\}, \Omega \subseteq \{1,2,...,n\} = C_n^1 + C_n^2 + ... + C_n^n = 2^n$.

В случае a) значение k находится из решения задачи обучения (оптимизации модели) или задается экспертом.

В данной статьи, в отличие от [4, 5], при построении системы опорных множеств Ω_A используются не все комбинации, а только $C_n^{n_0}$ $(n_0 = 1; n_0 = 2; ...; n_0 = k; (k < n))$ комбинации признаков. Здесь $n_0 (n_0 < n)$ заранее определяется с учетом требуемых значений качества и надежности распознавания при заданном объеме (число признаков и объектов) эталонной выборки, т.е. $n_0 = f(\varepsilon, \eta, n, m)$, где ε - вероятность ошибки, η - надежность распознавания, n - количество признаков, m - количество объектов. Это приводит к резкому сокращению системы опорных множеств Ω_A и позволяет при распознавании объектов использовать комбинации признаков входящих в $\Omega_A = C_n^{n_0}$.

Пусть для образов $K = K_1,...,K_l$ выполняется $K_i \cap K_j = \phi, \forall i \neq \forall j$. Обозначим через T_{nml} - эталонную выборку, $T^*_{nm_1}$ - контрольную выборку, где n -количество признаков, m - количество объектов, l - количество классов, m_1 - количество объектов контрольной выборки.

Требуется, используя эталонную выборку T_{nml}, найти предельно-допустимую размерность пространства комбинаций признаков n_0 , построить систему опорных множеств Ω_A из комбинаций признаков $C_n^{n_0}$ и определить на этом множестве решающие правила, обеспечивающие требуемые качество и надежность распознавания.

Формирование системы опорных множеств. В [4,5] системы опорных множеств из комбинаций признаков в первом случае определяется как $\Omega_A = C_n^k$ и во втором случае как $\Omega_A = C_n^1 + C_n^2 + ... + C_n^n = 2^n$. Если учесть, что каждый объект $S_i \in T^*_{nm_1}$ сопоставляется с каждым объектом $S_j \in T_{nml}$ с помощью $C_n^i (i = \overline{1,k}$ либо $i = \overline{1,n})$ на основе правила

$$d(S_i, S_j) = \begin{cases} 1, если \ \tilde{\omega}S_i = \tilde{\omega}S_j \\ 0, если \ \tilde{\omega}S_i \neq \tilde{\omega}S_j \end{cases} , (2)$$

тогда результаты сопоставления будут соответствовать одному из вариантов 2^i, где $\tilde{\omega}S_i$ и $\tilde{\omega}S_j$ называются ω-частью объектов S_i и S_j соответственно. Следовательно, из [4,5] следует, что если учесть в первом случае $\Omega_A = C_n^k$ и 2^i, во втором случае $\Omega_A = C_n^1 + C_n^2 + ... + C_n^n = 2^n$ и 2^i, тогда число всевозможных подмножеств $\{1,2,...,n\}$ длины i и результаты вариантов сопостовления объектов S_i и S_j будут $N = 2^i C_n^i,\ i = \overline{1,k}$ и $N = 2^i C_n^i,\ i = \overline{1,n}$ соответственно.

Если из эталонной выборки T_{nml} определена система опорных множеств $\Omega_A = C_n^{n_0}$ и всевозможные варианты результатов 2^{n_0} сопоставления объектов $S_i \in T^*_{nm_1}$ с объектами $S_j \in T_{nml}$, то число всевозможных решающих правил составит величину, меньшую N, где

$$N = 2^{n_0} C_n^{n_0} . \qquad (3)$$

Предположим, что до обучения заранее заданы требуемые значения вероятности ошибки ε и надежность распознавания η, а также количество признаков n и объектов m. Тогда из (1) можно получить функциональную зависимость

$$ln\, N = \varepsilon m + ln\, \eta . \qquad (4)$$

Для того чтобы найти n_0 логарифмируем (3)

$$\ln N = \ln 2^{n_0} + \ln C_n^{n_0} . \qquad (5)$$

Используя $C_m^n \le \frac{m^n}{2^n}$ и учитывая (6) из соотношения (5) получим

$$\ln N = n_0 \ln n \qquad (6)$$

Если полученное значение (6) подставить в соотношение (4), то можно получить конкретную функциональную зависимость для n_0

$$n_0 = \frac{\varepsilon m + \ln \eta}{\ln n} . \qquad (7)$$

Найденное по этой зависимости значение n_0 гарантирует заданную вероятность ошибки ε с надежностью $(1-\eta)$ при фиксированных m, n, η.

Если зафиксировать η, m, n, ε, то из соотношения (7) можно найти предельные значения размерности n_0 пространства комбинаций признаков, удовлетворяющие заданной вероятности ошибки ε при распознавании новых объектов (Таблица 1). Таблица 1.

$\eta = 0.95$, $m = 100$, $n = 50$.					
ε	0,04	0,07	0,12	0,15	0,19
n_0	1	2	3	4	5

Анализ приведенных в таблице 1 данных показывает, что увеличение вероятности ошибки ε при распознавании новых объектов, приводит к увеличению размерности n_0 пространства комбинаций признаков.

Если зафиксировать η, n_0, ε, n, то из соотношения (7) можно найти требуемое количество объектов $m = \frac{n_0 \ln n + \ln \eta}{\varepsilon}$, удовлетворяющих заданной вероятности ошибки ε при распознавании новых объектов (Таблица 2).

Таблица 2

$\eta = 0.95$, $n = 20$.					
n	1	2	3	4	5
ε	0,02	0,03	0,04	0,05	0,06
m	294	297	298	298	299

Алгоритм. Рассмотрим алгоритм, позволящий формировать систему опорных множеств $\Omega_A = C_n^{n_0}$, определяющих в этом множестве решающие правила $F(S_j), (S_j \in T^*_{nm_1})$, которые обеспечивают требуемое качество и надежность распознавания.

Алгоритмом сначала из эталонной выборки T_{nml} формируется система опорных множеств $\Omega_A = C_n^{n_0}$, далее в этом множестве каждый объект $S_j \in T^*_{nm_1}, j = \overline{1, m_1}$ сопоставляется с объектами $S_1, ..., S_{m_j} \in K_j, j = \overline{1, l}$ и в результате суммарного голосования (суммарной степени близости распознаваемого объекта S_i к классу K_j) каждый объект $S_j \in T^*_{nm_1}, j = \overline{1, m_1}$ классифицируется, одному из заранее заданных классов $K_j \in T_{nml}, j = \overline{1, l}$. При этом в результате классификации новых объектов $S_j \in T^*_{nm_1}, j = \overline{1, m_1}$ обеспечиваются требуемые ε и η с учетом заданных n, m.

В отличие от алгоритмов, приведенных в [4, 5], в данном алгоритме:

- при заданных ε и η, а также при фиксированных n и m определяется n_0 в виде (7).

- формируется система опорных множеств из заданных n признаков по n_0, т.е. $\Omega = C_n^{n_0}$ $(n_0 \le n)$;

- вычисляется $\Omega_A = C_n^{n_0}$ сумма голосов схожести $\Gamma(S_j) \in T^*_{nm_1}, j = \overline{1, m_1}$ для каждого нового объекта $S_j \in T^*_{nm_1}, j = \overline{1, m_1}$;

- классифицируются объекты $\Gamma(S_j) \in T^*_{nm_1}, j = \overline{1, m_1}$ по сумме результатов голосования, одному из заданных классов $K_j \in T_{nml}, j = \overline{1, l}$.

- резко снижается объем вычислений на компьютере, так как $C_n^{n_0} \le C_n^k$ ($n_0 < k$) и $C_n^{n_0} \le C_n^1 + C_n^2 + \ldots + C_n^n$ ($n_0 < n$).

Отсюда видно, что в данном алгоритме сильно сокращаются системы опорных множеств Ω_A. Одновременно, за счет сокращения системы опорных множеств уменьшается множество, что приводит зачастую к снижению надежности распознавания в целом.

Алгоритм включает в себя следующие основные этапы:

1. В оперативную память заносятся объекты $S_1, S_2, \ldots, S_m$, их признаки $x_1, x_2, \ldots, x_n$ и классы $K_1, K_2, \ldots, K_l$ в виде эталонной выборки T_{nml}.

2. В оперативную память заносятся объекты $S_1, S_2, \ldots, S_{m_1}$ и их признаки $x_1, x_2, \ldots, x_n$ в виде контрольной выборки $T^*_{nm_1}$.

3.Определяется предельное значение допустимой размерности n_0 в виде (7), с учетом заданных значениях ε и η, а также при фиксированных n и m.

4.Формируются Ω_A из заданных n признаков по n_0, т.е. $\Omega_A = C_n^{n_0}$.

5. $i = 1$. Из оперативной памяти отбирается объект $S_i \in T^*_{nm_1}$.

6. $j = 1$. Отбирается объект $S_j \in T_{nml}$.

7. Сопоставляется объект $S_i \in T^*_{nm_1}$ с объектом $S_j \in T_{nml}$ по правилу (2).

8. Ставится объекту $S_j \in T_{nml}$ 0 (ноль) либо 1 (единица) в зависимости от результатов схожести, который производиться в 7-шаге алгоритма.

9. $j = j + 1$. Если $j \le m$, то алгоритм переходит к шагу 6, в противном случае к шагу 10.

10. Формируется новая таблица T_{Ω_A}, полученных по системе опорных множеств $\Omega_A = C_n^{n_0}$.

11. $p = 1$. Выделяется класс K_p и относительно этого класса вычисляется сумма голосов $\Gamma_\Sigma(K_p)$.

12. $\Gamma_\Sigma(K_p) = 0$.

13. $j = 1$. Выделяется объект $S_j \in K_p$.

14. $\Gamma_{\Omega_A}(S_j) = 0$.

15. $k = 1$.

16. Проверяется условие:

а) Если $d(\tilde{\omega} S_{ik}, \tilde{\omega} S_{jk}) = 1$, то $\Gamma_{\Omega_A}(S_j) = \Gamma_{\Omega_A}(S_j) + 1$, и алгоритм переходит к шагу 17.

б) Если $d(\tilde{\omega} S_{ik}, \tilde{\omega} S_{jk}) = 0$, то алгоритм переходит к шагу 17.

17. $k = k + 1$. Если $k \le t (t = C_n^{n_0})$, то алгоритм переходит к шагу 15, в противном случае к шагу 18.

18. $j = j + 1$. Если $j \le m_i$, то алгоритм переходит к шагу 13, в противном случае к шагу 19.

19. $p = p + 1$. Если $p \le l$, то алгоритм переходит к шагу 11, в противном случае к шагу 20.

20. Вычисляются сумма голосов схожести распознаваемого объекта S_i к классу K_p:

$$\Gamma_\Sigma(S_i \in K_1) = \sum_{j=1}^{m_1} \sum_{k=1}^{t} \Gamma_{\Omega_A}(S_{jk})$$

$$\Gamma_\Sigma(S_i \in K_2) = \sum_{j=1}^{m_2} \sum_{k=1}^{t} \Gamma_{\Omega_A}(S_{jk})$$

....

$$\Gamma_\Sigma(S_i \in K_l) = \sum_{j=1}^{m_l} \sum_{k=1}^{t} \Gamma_{\Omega_A}(S_{jk})$$

21. Для распознавания объекта $S_i \in T^*_{nm_1}$ используется решающее правило

$$F(S_i): \begin{cases} S_i \in K_p, \text{ если } \max\limits_{1 \le p \le l} \{\Gamma_\Sigma(S_i \in K_1), \Gamma_\Sigma(S_i \in K_2), \ldots, \Gamma_\Sigma(S_i \in K_l)\} \\ S_i \notin K_1, S_i \notin K_2, \ldots, S_i \notin K_l \text{ если } \Gamma_\Sigma(S_i \in K_1) = \Gamma_\Sigma(S_i \in K_2) = \ldots = \Gamma_\Sigma(S_i \in K_l) \end{cases}$$

22. $i = i + 1$. Если $i \le m_1$, то алгоритм переходит к шагу 5, в противном случае к шагу 23.

23. Вывод результатов, относящих объект $S_i \in T^*_{nm_1}$ по сумме голосований за классы, в один

из классов $K_1, K_2, ..., K_l \in T_{nml}$, или указывающее для объекта $S_i \in T^*_{nm_1}$ отказ от распознавания.

Программный комплекс. Создан комплекс программ на основе разработанного алгоритма. Общий вид интерфейсного окна программы имеет следующий вид (Рис. 1).

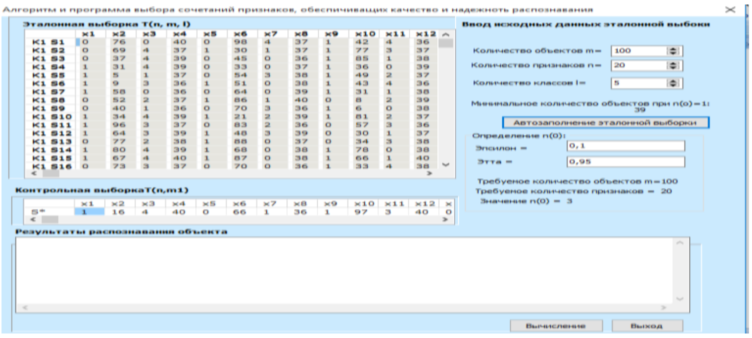

Рис. 1. Общий вид интерфейсного окна.

Модуль вычисления n_0, системы опорных множеств $\Omega_A = C_n^{n_0}$ и распознавания объектов по $\Omega_A = C_n^{n_0}$ представлен на рис. 2.

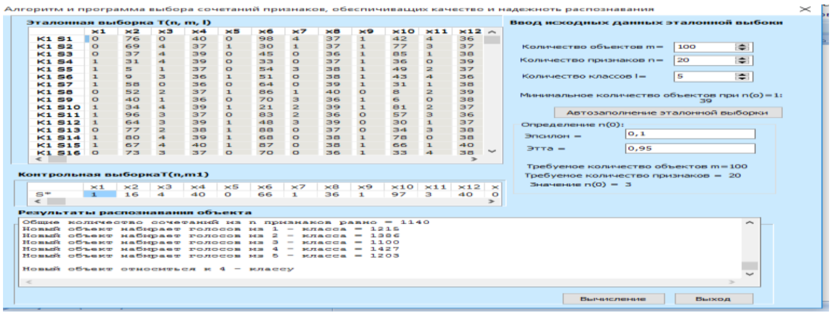

Рис. 2. Результаты вычислительного эксперимента.

Проведено испытание по оценке работоспособности и эффективности предложенного алгоритма и программного комплекса применительно к распознаванию образов. Полученные результаты подтверждают то, что разработанный алгоритм и программный комплекс применим для решения практических задач распознавания объектов, касающихся медицинской, технической, археологической, гидрогеологической, сейсмологической, биологической и геологической сферы.

Список литературы:

1. Бекмуродов К.А., Васильев В.И., Бекмуратов Д.К. Нахождение предельно-допустимых значений размерности признаковых пространств из обучающей выборки. //Академия Наук Республики Узбекистан. Институт математики и информационных технологий. Современное состояние и перспективы развития информационных технологий. Том 2. Ташкент, 2011. 309-312 с.

2. Бекмуродов Қ.А., Бекмуратов Д.Қ. Последовательный выбор признаков, обладающих требуемой разделяющей силой. XI - Международная научно-практическая конференция. "Научные перспективы XXI века. Достижения и перспективы нового столетия". Ежемесячный научный журнал №4(11) / 2015, часть 4. Россия, г. Новосибирск, 22-23.05, 2015 г. 9-13 с. ISSN 34567-1769.

3. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). М.:Наука, 1974. - 412 с.