

**AVTOMATİK TƏSNİFAT MƏSƏLƏLƏRİNİN HƏLLİ
ÜÇÜN YENİ ALQORİTM VƏ ONUN HESABLANMA
MÜRƏKKƏBLİYİNİN TƏDQIQI**

R.İ.DAVUDOVA
AMEA Kibernetika İnstitutu

Təqdim edilən işdə ədədi əlamətləri ilə təsvir edilən obyektlərin avtomatik təsnifatının qurulması üçün sinfin "orta"sı adlandırılan saxta obyektədən bu sinfə daxil olan digər obyektlərin orta kvadratik meyilləri cəminin minimumlaşdırılmasına əsaslanan yeni A1 alqoritmi təklif edilmiş, bu alqoritmin və eyni zamanda bir neçə obyektin bir sinfdən digər sinfə keçirilməsi ilə məlum təsnifatı dəqiqləşdirməyə imkan verən yeni A2 alqoritminin hesablanma mürəkkəbliyi tədqiq edilmiş və onların polinomial alqoritmlər sinfindən olduğu isbat edilmişdir.

Hazırda çoxölçülü avtomatik təsnifat məsələlərinin həlli üçün müxtəlif alqoritmlər işlənib hazırlanmışdır. Bu alqoritmlər obyektlərin çoxölçülü əlamətlər fəzasında müxtəlif paylanma meyarlarına əsaslanır [1-7].

Çoxölçülü əlamətlər fəzasında verilmiş obyektlərin siniflər arasında yenidən bölgüsünün aparılması optimal təsnifatın qurulması məsələlərində mühüm addımdır. [3,4]-də bütün siniflər üzrə "sinfin ortası (mərkəzi)" adlandırılan saxta elementdən bu sinfə daxil olan digər obyektlərə qədər olan orta kvadratik meyillərin cəminin minimumlaşdırılmasına əsaslanan yeni həlledici qayda təklif edilmiş və bu qayda təkmilləşdirilərək, ilk dəfə olaraq, bir neçə obyektin eyni zamanda bir sinfdən digərinə keçirilməsinə imkan verən ümumiləşmiş həlledici qayda işlənib hazırlanmışdır. [5]-də təklif edilən alqoritm obyektlərin bir sinfdən digər sinfə çoxluq halında keçirilməsi ilə məlum təsnifatı dəqiqləşdirməyə imkan verir. Təqdim edilən işdə isə bu həlledici qaydalar əsasında obyektlərin siniflər arasında bir-bir yenidən bölüşdürülməsi ilə məlum təsnifatı dəqiqləşdirməyə xidmət edən hesablama alqoritmi təklif edilmişdir.

Tutaq ki, $X = \{x_1, x_2, \dots, x_m\}$ obyektləri, onların n -ölçülü $I(x_1), I(x_2), \dots, I(x_m)$ təsvirləri məlumdur. Mümkün $I(x_1), I(x_2), \dots, I(x_m)$ təsvirləri və təsnifatın keyfiyyət funksionalının optimallaşdırılması əsasında $x_i \in X, i = \overline{1, m}$ obyektini üçün

$$P_j(x_i) - "x_i \in K_j", j = \overline{1, l}$$

predikatının hesablanması tələb edilir. Başqa sözlə, elə A alqoritmi qurmaq lazımdır ki,

$$A(I(x_1), I(x_2), \dots, I(x_m)) = \{\alpha_j^A(x_i)_{m \times l}\},$$

$$\alpha_j^A(x_i) = P_j(x_i), \quad i = \overline{1, m}, \quad j = \overline{1, l},$$

burada $\alpha_j^A(x_i) \in \{0, 1, \Delta\}$. Düzbucaqlı $T = \{x_{ij}\}_{m \times n}$ matrisini quraq. Belə ki, onun x_{ij} elementi i -ci obyektin j -ci əlaməti olsun. Başqa sözlə, T matrisinin p -ci sətiri X çoxluğunun p -ci obyektinə uyğun olsun.

Tutaq ki, bu və ya digər üsulun tətbiqi ilə müəyyən $\tilde{K} = (\tilde{K}_1, \tilde{K}_2, \dots, \tilde{K}_{\tilde{l}})$ təsnifatı qurulmuşdur, lakin o, təsnifatın keyfiyyətini səciyyələndirən optimallıq meyarını ödəmir. Belə optimallıq meyarı müxtəlif şəkillərdə seçilə bilər. Bu işdə keyfiyyət meyarı qismində siniflər üzrə orta (mərkəzi) obyekt olaraq seçilmiş saxta elementdən digər obyektlərə qədər orta kvadratik meyillərin cəmi çıxış edir və obyektlərin yenidən paylanması onun minimumlaşdırılması əsasında aparılır:

$$\Phi = \sum_{j=1}^{\tilde{l}} \sum_{x \in \tilde{K}_j} \|x - \bar{x}_j\|^2, \quad (1)$$

burada \tilde{l} \tilde{K} təsnifatında mövcud siniflərin sayı, \bar{x}_j \tilde{K}_j sinfi üzrə saxta olaraq daxil edilmiş mərkəzi obyektidir və

$$\bar{x}_j = \frac{1}{m_j} \sum_{p=1}^{m_j} x_p, \quad j = \overline{1, \tilde{l}}. \quad (2)$$

Belə ki, burada m_j \tilde{K}_j sinfinə daxil olan obyektlərin sayını göstərir. (1) münasibətindəki $\|\bullet\|$ işarəsi Euklid mənada normanı göstərir və

$$\|x\| = \left(\sum_{i=1}^n x_i^2 \right)^{1/2}$$

kimi müəyyən edilir.

Əvvəlki bölgüyə nisbətən daha optimal, başqa sözlə, daha dəqiq təsnifatın qurulması obyektlərin siniflər arasında yenidən bölüşdürülməsi ilə aparılmalıdır.

Tutaq ki, [2-4]-də göstərildiyi kimi, \tilde{K}_i və \tilde{K}_j , $i \neq j$ siniflərinə baxılır və x_i obyektinin \tilde{K}_i sinfindən olduğu qəbul edilir, yəni $x_i \in \tilde{K}_i$.

Teorem 1.[4]: $x_i \in \tilde{K}_i$ obyektinin \tilde{K}_j -sinfinə aid edilməsi üçün zəruri və kafi şərt

$$\frac{m_j}{m_j + 1} \|x_i - \bar{x}_j\|^2 < \frac{m_i}{m_i - 1} \|x_i - \bar{x}_i\|^2 \quad (3)$$

münasibətinin ödənilməsidir.

(3) bir obyektin bir sinifdən digərinə keçirilməsinə imkan verən həlledici qaydanın ifadəsidir.

İndi isə obyektlərin ilkin $\tilde{K} = (\tilde{K}_1, \tilde{K}_2, \dots, \tilde{K}_{\tilde{l}})$ məlum təsnifatını (3) həlledici qaydası əsasında dəqiqləşdirmək üçün yeni A1 alqoritmini verək. Qeyd edək ki, bu alqoritm işə başlamazdan əvvəl siniflərin \tilde{l} sayı və müəyyən ilkin $\tilde{K}^0 = (\tilde{K}_1^0, \tilde{K}_2^0, \dots, \tilde{K}_{\tilde{l}}^0)$ təsnifatı məlum olmalıdır.

Alqoritmlərin təsvirinə keçməzdən əvvəl, istifadə edilən bəzi işarələmələri qəbul edək:

$T = [x_{pq}]$: $p = \overline{1, m}$, $q = \overline{1, n}$; x_{pq} - T cədvəlinin p -ci sətir və q -cü sütunun kəsişməsində duran element və ya başqa cür ifadə etsək, p -ci obyektin q -cü əlamətinin ədədi xarakteristikası; \tilde{l} - əvvəlcədən məlum olan təsnifatı təşkil edən siniflərin sayı; k - iterasiyanın nömrəsi; i, j - uyğun olaraq, \tilde{K}_i, \tilde{K}_j siniflərinin nömrəsi (indeksi); $\bar{x}_p(k)$ k -cı iterasiyada \tilde{K}_p sinfinin saxta mərkəzi; $m_p(k)$, $p = 1, 2, \dots, \tilde{l}$ k -cı iterasiyada \tilde{K}_p sinfini təşkil edən obyektlərin sayıdır;

A1 alqoritmi:

Addım 1: $k = 0$ qəbul edilir.

Addım 2: $k = k + 1$

$$m_p(k) = m_p(k-1), p = \overline{1, \tilde{l}};$$

$$\tilde{K}^k = (\tilde{K}_1^k, \tilde{K}_2^k, \dots, \tilde{K}_{\tilde{l}}^k) = (\tilde{K}_1^{k-1}, \tilde{K}_2^{k-1}, \dots, \tilde{K}_{\tilde{l}}^{k-1}) = \tilde{K}^{k-1}$$

götürülərək, addım 3-ə keçməli.

Addım 3: $\tilde{K}^k = (\tilde{K}_1^k, \tilde{K}_2^k, \dots, \tilde{K}_{\tilde{l}}^k)$ təsnifatı üçün

$$\bar{x}_p(k) = \frac{1}{m_p(k)} \sum_{x_q \in \tilde{K}_p(k)} x_q, \quad p = \overline{1, \tilde{l}} \quad (4)$$

saxta mərkəzləri hesablanır.

Addım 4: $i = 1$ götürülür (i - o sinfin nömrəsidir ki, həmin sinifdən obyektin daşınması nəzərdə tutulur).

Addım 5: $j = 1$ götürülür (j - o sinfin nömrəsidir ki, həmin sinifə obyektin daşınması nəzərdə tutulur).

Addım 6: Əgər $i = j$ olarsa, onda addım 11-ə, əks halda isə addım 8-ə keçməli.

Addım 7: $\tilde{K}_i(k)$ və $\tilde{K}_j(k)$ sinifləri üçün saxta orta element (obyekt) hesablanır:

$$\bar{x}_i(k) = \frac{1}{m_i(k)} \sum_{x_q \in \tilde{K}_i(k)} x_q, \quad \bar{x}_j(k) = \frac{1}{m_j(k)} \sum_{x_p \in \tilde{K}_j(k)} x_p \quad (5)$$

Addım 8: $s = 1$ qəbul edilir.

Addım 9: $x_s \in \tilde{K}_i(k)$ obyektı üçün

$$\frac{m_j(k)}{m_j(k)+1} \|x_s - \bar{x}_j(k)\|^2 < \frac{m_i(k)}{m_i(k)-1} \|x_s - \bar{x}_i(k)\|^2 \quad (6)$$

şerti yoxlanılır.

a) Əgər şərt ödənilirsə, x_s obyektı $\tilde{K}_i(k)$ sinfindən $\tilde{K}_j(k)$ sinfinə keçirilir və aparılmış yeni bölgüdən sonra $\tilde{K}_i(k)$ və $\tilde{K}_j(k)$ siniflərində olan obyektlerin sayı, uyğun olaraq,

$$m_i(k) = m_i(k) - 1$$

$$m_j(k) = m_j(k) + 1$$

düsturları ilə hesablanır. Yeni orta elementlər (5) münasibətlərinə görə korrektə edilir. Addım 10-a keçməli.

b) Əgər (6) şərti ödənilmirsə, onda addım 10-a keçməli.

c) (6) münasibətində bərabərlik halı alınarsa, seçim ixtiyari aparılır. Başqa sözlə ifadə etsək, obyektler çoxluğu $\tilde{K}_i(k)$ və $\tilde{K}_j(k)$ siniflərindən istənilən birinə daxil edilə bilər.

Addım 10: $s = s + 1$ götürülür və $s \leq m_i(k)$ şərti yoxlanılır. Əgər şərt ödənilirsə, onda addım 9-a, əks halda isə addım 11-ə keçməli.

Addım 11: $j = j + 1$ qəbul edilir. Əgər $j \leq \tilde{l}$ şərti ödənilirsə, addım 6-ya, əks yalda isə addım 12-yə keçməli.

Addım 12: $i = i + 1$ qəbul edilir və $i \leq \tilde{l}$ şərti yoxlanılır. Şərtin ödənilmədiyi halda addım 5-ə, əks halda isə addım 13-ə keçilir.

Addım 13: Alqoritmin yığılması üçün

$$\bar{x}_p(k+1) = \bar{x}_p(k), \forall p = \overline{1, \tilde{l}} \quad (7)$$

münasibəti yoxlanılır. Bu münasibət doğru olarsa, proses dayandırılır. Münasibətin ödənilmədiyi halda isə addım 14-ə keçməli.

Addım 14: $m_p(k) \geq 1, p = \overline{1, \tilde{l}}$ şərti yoxlanılır. Əgər şərt ödənilirsə, addım 2-yə, əks halda isə addım 15-ə keçməli.

Addım 15 : $\tilde{l} = \tilde{l} - 1$ qəbul edilir və addım 2-yə keçməli.

Alqoritmin təsvirindən də görüldüyü kimi, prosesin gedşində siniflərin sayı dəyişə bilər. Belə ki, alqoritm prosesində bəzi siniflər boşala bilər və belə olduqda, həmin siniflər ləğv edilir.

Qeyd edək ki, yuxarıdakı münasibətlərdə istifadə edilən $x_p, p = \overline{1, m}$ və $\bar{x}_j(k), j = \overline{1, \tilde{l}}$ kəmiyyətləri n -ölçülü vektor kəmiyyətlərdir.

İndi isə **A1** və [5]-də təklif edilən **A2** alqoritmlərinin hesablanma mürəkkəbliyini, başqa sözlə, bu alqoritmlərin bir iterasiyasının icrası zamanı tələb edilən əməliyyatların maksimal sayının qiymətləndirməsini verək. Burada m obyektlerin sayı, n obyektlerin ədədi xarakteristikalarının (əlamətlərinin) sayı, \tilde{l} isə təsnifatı təşkil edən siniflərin sayıdır.

Qeyd edək ki, bu alqoritmlərin hər ikisində obyektlərin yenidən bölüşdürülməsi cüt-cüt götürülmüş siniflər arasında $i \neq j$ olmaqla aparılır. \tilde{l} sayda sinfin bütün cüt-cüt qruplaşmalarının sayı

$$A_{\tilde{l}}^2 = \frac{\tilde{l}!}{(\tilde{l}-2)!} = \tilde{l}(\tilde{l}-1)$$

münasibəti ilə təyin edilir. $i \neq j$ şərtini ödəyən qruplaşmaların sayı isə

$$\tilde{l}^* = A_{\tilde{l}}^2 - \tilde{l} = \tilde{l}(\tilde{l}-1) - \tilde{l} = \tilde{l}^2 - 2\tilde{l}$$

olar.

Əvvəlcə **A1** alqoritminin ayrı-ayrı mühüm addımları üçün qiymətləndirmələri verək:

- \tilde{l} sayda sinif üçün orta obyektin hesablanması: $O(m \cdot n \cdot \tilde{l})$;
- (6) şərtinin yoxlanılması: $O(m \cdot n \cdot \tilde{l}^2)$;
- Alqoritmın dayandırılması şərti: $O(n \cdot \tilde{l})$.

Deməli, **A1** alqoritminin hesablanma mürəkkəbliyinin qiymətləndirilməsi üçün

$$T(\mathbf{A1}) = O(m \cdot n \cdot \tilde{l}) + O(m \cdot n \cdot \tilde{l}^2) + O(n \cdot \tilde{l}) = O(m \cdot n \cdot \tilde{l}^2) \quad (8)$$

münasibətini aldıq.

Teorem 2: Məlum təsnifatı dəqiqləşdirməyə imkan verən **A1** alqoritminin bir iterasiyası ən çoxu $O(m \cdot n \cdot \tilde{l}^2)$ sayda əməlin icrasını tələb edir və deməli, bu alqoritm *P* effektiv alqoritmlər ailəsindəndir.

İndi isə **A2** alqoritminin hesablanma mürəkkəbliyini qiymətləndirək:

- \tilde{l} sayda sinif üzrə orta obyektin hesablanması: $O(m \cdot n \cdot \tilde{l})$;
- \tilde{l} sayda sinifə daxil olan obyektlərin orta obyektədən məsafələrinin hesablanması: $O(m \cdot n \cdot \tilde{l})$;
- \tilde{l} sayda sinifə daxil olan obyektlərin orta obyektədən məsafələrinin nizamlanması: $O(m^2 \cdot \tilde{l})$;
- obyektlər çoxluğunun siniflər arasında yenidən bölüşdürülməsi: $O(m \cdot n \cdot \tilde{l}^*) = O(m \cdot n \cdot \tilde{l}^2)$;
- Alqoritmın dayandırılması şərti: $O(n \cdot \tilde{l})$.
- Onda **A2** alqoritminin hesablanma mürəkkəbliyinin yuxarı sərhədi üçün
- $T(\mathbf{A2}) = O(m \cdot n \cdot \tilde{l}) + O(m^2 \cdot \tilde{l}) + O(m \cdot n \cdot \tilde{l}^2) + O(n \cdot \tilde{l}) =$
 $= O(m^2 \cdot n \cdot \tilde{l}^2) \quad (9)$

qiymətləndirməsini alırıq.

Beləliklə, aşağıdakı teoremi isbat etdik:

Teorem 3: Məlum təsnifatı dəqiqləşdirməyə imkan verən **A2** alqoritminin bir iterasiyası gedişində icra edilən əməllərin maksimal sayı

$O(m^2 \cdot n \cdot \tilde{l}^2)$ tərtibdəndir və bu alqoritm P polinomial alqoritmlər sinfindəndir.

ƏDƏBİYYAT

1. Журавлев Ю.И., Никифоров В.В. Алгоритмы распознавания, основанные на вычислении оценок. // Кибернетика, 1971, №3, с.1-11.
2. Зенкин А.И., Керимов А.К. Задача построения оптимальных классификаций динамических объектов. М.: ВЦ АН СССР, 1985 г., с.20, препринт.
3. Керимов А.К. Построение оптимальных классификаций динамических объектов и прогнозирование состояний сложных систем. Автореф. дисс. на соис. уч.ст. к.ф.-м.н., Москва: 1985, с.18.
4. Керимов А.К. Новое решающее правило в задаче автоматической классификации // Изв. НАН Азербайджана, сер. физ.-мат. и тех. наук, т. XXV, № 3, 2005, с.95-98.
5. Давудова Р.И. Об одном алгоритме построения оптимальной классификации // Изв. НАН Азербайджана, сер. физ.-мат. и тех. наук, т. XXV, № 3, 2005, с.227-230.
6. Kərimov Ə.K., Davudova R.İ. Optimal təsnifatın qurulması məsələsinə yeni yanaşma / AMEA müx. üzvü İ.T.Məmmədovun anadan olm.-nın 50 illiyinə həsr edilmiş beynəlxalq elmi konfrans, 11-13 may 2005, с.112.
7. Давудова Р.И. Исследование вычислительной сложности алгоритма оптимальной классификации, базирующихся на генетических алгоритмах / Сборник докладов межд. конф. по мягким вычислениям и измерениям (SCM'2007), Санкт-Петербург, 25-27 июня 2007 г, т.1, с.283-285.

НОВЫЙ АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ И ИССЛЕДОВАНИЕ ЕГО ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ

Р.И.ДАВУДОВА
РЕЗЮМЕ

Рассматривается задача автоматической классификации объектов, описываемых количественными признаками. Для улучшения начальной классификации предложен алгоритм \mathcal{A} , основанный на минимизации суммы среднеквадратичных отклонений объектов от фиктивного “среднего” объекта. Исследованы вычислительные сложности алгоритма \mathcal{A} и алгоритма \mathcal{B} , который уточняет начальную классификацию перетаскиванием набора объектов из одного класса в другой. Доказано, что оба алгоритма имеют полиномиальную сложность.

NEW ALGORITHM OF THE SOLUTION OF THE AUTOMATIC CLASSIFICATION PROBLEM AND RESEARCH OF ITS COMPUTING COMPLEXITY

R.I.DAVUDOVA

SUMMARY

It is considered the automatic classification problem of the objects described by quantitative attributes. For improvement of initial classification algorithm \mathcal{A} based on minimization of the sum mean-square deviation of objects from fictitious “average” object is offered. Computing complexity of algorithm \mathcal{A} and of algorithm \mathcal{B} which specifies initial classification by reference of the objects set from one class in another is investigated. It is shown, that the both algorithms have polynomial complexity.