

3-7-2018

## MODERN ALGORITHMS OF RECOGNIZING SYMBOLS

Z B. Mingliqulov

*Tashkent University of Information Technologies named after Muhammad al-Kharezmi*

D M. Sotvoldiev

*Fergana branch of the Tashkent University of Information Technologies named after Muhammad al-Kharezmi*

M U. Norinov

*Fergana branch of the Tashkent University of Information Technologies named after Muhammad al-Kharezmi*

Follow this and additional works at: <https://uzjournals.edu.uz/ferpi>

---

### Recommended Citation

Mingliqulov, Z B.; Sotvoldiev, D M.; and Norinov, M U. (2018) "MODERN ALGORITHMS OF RECOGNIZING SYMBOLS," *Scientific-technical journal*: Vol. 1 : Iss. 1 , Article 8.

Available at: <https://uzjournals.edu.uz/ferpi/vol1/iss1/8>

This Article is brought to you for free and open access by 2030 Uzbekistan Research Online. It has been accepted for inclusion in Scientific-technical journal by an authorized editor of 2030 Uzbekistan Research Online. For more information, please contact [sh.erkinov@edu.uz](mailto:sh.erkinov@edu.uz).

УДК 519.681.5

**MODERN ALGORITHMS OF RECOGNIZING SYMBOLS****Z.B. Mingliqulov<sup>1</sup>, D.M. Sotvoldiev<sup>2</sup>, M.U. Norinov<sup>2</sup>**<sup>1</sup>Tashkent University of Information Technologies named after Muhammad al-Kharezmi<sup>2</sup>Fergana branch of the Tashkent University of Information Technologies named after Muhammad al-Kharezmi**СОВРЕМЕННЫЕ АЛГОРИТМЫ РАСПОЗНАВАНИЯ ИЕРАРХИЙ****З.Б. Мингликулов<sup>1</sup>, Д.М. Сотволдиев<sup>2</sup>, М.У. Норинов<sup>2</sup>**<sup>1</sup>Ташкентского университета информационных технологий имени

Мухаммад аль-Харезми

<sup>2</sup>Ферганский филиал Ташкентского университета информационных технологий имени Мухаммад аль-Харезми**ТИМСОЛЛАРНИ ТАНИБ ОЛИШНИНГ ЗАМОНАВИЙ АЛГОРИТМЛАРИ****З.Б. Мингликулов<sup>1</sup>, Д.М. Сотволдиев<sup>2</sup>, М.У. Норинов<sup>2</sup>**<sup>1</sup>Мухаммад ал Хоразмий номидаги Тошкент ахборот технологиялари университети,<sup>2</sup>Мухаммад ал Хоразмий номидаги Тошкент ахборот технологиялари университети Фарғона филиали

**Abstract.** In this work, we consider one of the tasks of data mining, the problem of classification. It uses algorithms of the logical model of fuzzy classification and algorithms of  $k$  nearest neighbors. In the process of building a fuzzy logic model, in some cases, some methods from the base of fuzzy rules were used by the method of mountain clustering. In addition, the article proposes a new method for constructing a base of fuzzy rules, using the method of fuzzy clustering. In addition, in the article, the problem of classifying the Iris flower was solved based on some classification algorithms and the results were compared.

**Keywords:** data mining, soft computing,  $k$ NN, optimization of parameters, fuzzy model, algorithm of fuzzy clustering, demassification, logic model of fuzzy classification.

**Аннотация.** В данной работе рассмотрена одна из задач интеллектуального анализа данных, задача классификации. В нем использованы алгоритмы логическая модель нечеткой классификации и алгоритмы  $k$  ближайших соседей. В процессе построение нечеткой логической модели в некоторых случаях использованы некоторые методы из базы нечетких правил методом горной кластеризации. Кроме этого, в статье предложен новый метод построения базы нечетких правил, с использованием метода нечеткой кластеризации. Так же, в статье задача классификации цветка Ирис решено на основе некоторых алгоритмов классификации и проведено сравнение результатов.

**Ключевые слова:** интеллектуальный анализ данных, мягкое вычисление,  $k$ NN, оптимизация параметров, неясная модель, алгоритм неясной кластеризации, денацификация, логическая модель неясной классификации.

**Аннотация.** Ушбу мақолада маълумотларни интеллектуал таҳлил қилиш масаласидан бири бўлган синфлаштириш масаласи кўриб ўтилган. Бунда норавиан синфлаштириш мантиқий моделини қуриш алгоритми ва  $k$  яқин қўшнилар алгоритмларидан фойдаланилди. Норавиан мантиқий модел қуриш жараёнида аксарият ҳолларда норавиан қоидалар базаси мавжуд бўлган бир қатор усуллардан фойдаланилади, жумладан кластерлашнинг тоғли усулдан. Ушбу усулдан фарқли равишда, ишда норавиан қоидалар базасини норавиан кластеризация усули ёрдамида қуришнинг янги

узули таклиф этилган. Шу билан бирга, ушбу мақолада ирис гулини синфлаштириш масаласи бир қатор алгоритмлар асосида ечилган ва уларнинг натижалари ўзаро солиштирилган.

**Таянч сўзлар:** маълумотларни интеллектуал таҳлили, юмшоқ ҳисоблаш, kNN, параметрларни оптималлаштириш, норавшан модел, норавшан кластерлаш алгоритми, фаззификация, дефаззификация, норавшан синфлаштириш мантиқий модели.

## Кириш

Табиий шароитда инсон томонидан ечилиши мумкин бўлган синфлаштириш масалаларини шартли равишда икки гуруҳга ажратишимиз мумкин [1].

1. Шартлари, белгилар қиймати аниқ, равшан, тўлиқ шакллантирилган, ҳолда тақдим этиллувчи масалалар.

2. Объект параметрлари тўла шакллантирилмаган, яъни параметрлари норавшанлик билан тақдим этиладиган масалалар.

Биринчи турдаги масалаларни ечишда анъанавий математик аппарат ва компьютер дастурларидан фойдаланиш самарали натижа бериши мумкин. Аммо, кейинги гуруҳ масалаларини ечишда одатда, анъанавий математик моделлаштириш аппарати самара бермаслиги мумкин ва бундай масалаларни ечишда юмшоқ ҳисоблаш (*Soft Computing*) технологиялари воситаларидан фойдаланиш мақсадга мувофиқ бўлади.

Синфлаштириш масаласи – бу, жорий объектни параметрларининг қийматларига боғлиқ ҳолда олдиндан маълум бўлган синфлардан бирига киритиш демакдир. Ушбу масалада ҳар бир объект ўзининг белгилари сонидан келиб чиққан ҳолда  $N$ -ўлчамли вектор кўринишида тавсифланади. Синфлаштириш масаласини ечишда кўплаб равшан ва норавшан алгоритмлардан фойдаланиш мумкин. Ушбу мақолада равшан алгоритмлар сирасидан  $k$  яқин қўшнилар алгоритмини ва норавшан синфлаштириш алгоритми сифатида норавшан мантиқий хулосалаш моделларини кўриб ўтамиз.

### 1. **K яқин қўшнилар алгоритми (kNN - k Nearest Neighbor)**

Ушбу алгоритм янги объектни олдиндан маълум бўлган объектлар билан яқинлик даражасига кўра синфлаштиришга қаратилган алгоритмдир. Яъни, объектга яқин бўлган  $K$ -та қўшни объектлар қайси синфга тегишли бўлса, ушбу объект ҳам шу синфга тегишли деб баҳоланади. Алгоритмнинг асосий масаласи сифатида яқин қўшнилар сонини кўрсатувчи  $K$ -сонни аниқлашдан иборатдир. Чунки ушбу коэффициентнинг қиймати синфлаштириш сифатига сезиларли даражада таъсир кўрсатади.

Олайлик,  $m$  объектдан иборат бўлган ўқув танланма жадвал кўринишида берилган. Ҳар бир объект қайсидир бир синфга тегишли. Янги объект қайси синфга тегишли эканлигини ушбу ўқув танланмадан фойдаланган ҳолда аниқлаш талаб этилади.

Алгоритмнинг дастлабки қадамида яқин қўшнилар сонини ифодаловчи  $K$  сони аниқланади. Агарда  $K=1$  қийматни олса, алгоритм ўз моҳиятини йўқотади ва у жорий объектни ўзига энг яқин бўлган объект қайси синфга тегишли бўлса, ўша синфга киритади. Агар жуда катта қиймат белгиланган тақдирда ҳам алгоритм синфлаштириш масаласини нотўғри ечиши мумкин.

Иккинчи қадамда жорий объектга энг яқин бўлган  $k$  объект аниқланади (қўшнилар кидирилади). Масофани ҳисоблаш функцияси куйидаги талабларга жавоб бериши керак:

$d(x,y) \geq 0$ ,  $d(x,y) = 0$  шарт бажарилади, қачонки  $x = y$  бўлганда;

$d(x,y) = d(y,x)$ ;

$d(x,z) \leq d(x,y) + d(y,z)$ , шарт бажарилади, қачонки  $x,y,z$  нуқталар бир-бирининг устида ётмаса.

Бу ерда  $X, Y, Z$  – солиштирилаётган объектларнинг белгилари векторлари.

Ушбу алгоритмда белгилар қиймати ҳақиқий сонлардан ташкил топган бўлса, объектлар яқинлиги Евклид метрикаси асосида аниқланади:

$$D_E = \sqrt{\sum_i^n (x_i - y_i)^2},$$

бунда  $n$  – белгилар сони.

Агар белгилар қиймати бинар кўринишда бўлса, Хеминг метрикасидан фойдаланилади ва объектлар орасидаги фарқ функцияси қуйидагича бўлади:

$$dd(x, y) = \begin{cases} 0, & x = y, \\ 1, & x \neq y. \end{cases}$$

Одатда, ушбу алгоритми қўллаш жараёнида белгилар қийматлари устида нормаллаштириш жараёни амалга оширилади. Нормаллаштиришнинг кўплаб усуллари ва алгоритмларини кузатиш мумкин. Кенг қўлланилаётган нормаллаштириш алгоритми сифатида минимаксли нормаллаштириш формуласини келтириш мумкин:

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)}.$$

бунда  $\min(X)$  ва  $\max(X)$   $X$  белгининг минимум ва максимум қийматлари ҳисобланади.

Синфлаштириш масаласини ечиш давомида баъзи ҳолларда объект белгиларининг муҳимлик даражаси – вазни ҳам муҳим ҳисобланади. Белгиларнинг ушбу муҳимлик даражаси эксперт-мутахассислар томонидан ёки берилган эталон жадвал (ўқув танланма) асосида аниқлаб олиниши мумкин.

Белгилар вазнларига кўра масофани аниқлаш формуласини қуйидаги тарзда ифодалаш мумкин:

$$D_E = \sqrt{\sum_i^n Z_i (x_i - y_i)^2}.$$

Бунда  $Z_i$  мос равишда  $i$  – белгининг муҳимлик коэффиценти ҳисобланади.

## 2. Норовшан синфлаштириш мантиқий моделини қуриш алгоритми

Норовшан мантиқ “*Soft Computing*”нинг бир компоненти ҳисобланиб, унда лингвистик фикрлар кўринишидаги эксперт билимларидан фойдаланиш имконияти мавжуд. Норовшан мантиқни эволюцион алгоритмлар билан бирлаштириш асосида қурилган норовшан модел бутунлай янги сифатни беради. Бундай норовшан моделларда табиий тилдаги билимлардан фойдаланиш имконияти мавжуд бўлади.

Умумий ҳолатда норовшан синфлаштириш модели норовшан қоидалар хулосалари ёрдамида қуйидагича тавсифланади [2-4]:

$$\bigcup_{p=1}^{k_j} \left( \bigcap_{i=1}^n x_i = a_{i,jp}, w_{jp} \text{ вазн билан} \right) \rightarrow y_j = f_j(x_1, x_2, \dots, x_n).$$

Бу ерда  $a_{i,jp}$  -  $jp$  қаторнинг  $x_i$  ўзгарувчисини баҳоловчи лингвистик терм.

$w_{jp}$  -  $jp$  - қоиданинг вазн коэффиценти.

$y_j = f_j(x_1, x_2, \dots, x_n)$  - норовшан қоиданинг чиқиши.

Бугунги кунда кўплаб тадқиқот ишларида норовшан мантиқий модел қуриш жараёнида мавжуд тоғли кластерлаш усулидан фойдаланилган. Шунингдек кўплаб амалий математик масалалар пакетида ҳам норовшан модел қуриш масаласи тоғли кластерлаш усули ёрдамида амалга оширилган [4,5].

Ўтказилган қатор тадқиқотлар шуни кўрсатдики, норовшан мантиқий модел қуришда тоғли кластерлаш усулидан фойдаланилганида қоидалар фақат бир хил термларга кўра

кластерлаштирилар экан. Ушбу ишда норавшан қоида хулосалари ёрдамида Сугенонинг ночизикли кўринишдаги моделини норавшан кластерлаштириш усули ёрдамида қуриш ва модел параметрларини арилар колонияси алгоритми асосида сошлаш масаласи кўриб ўтилади.

Норавшан қоидалар хулосалари ёрдамида тавсифланувчи сушт шаклланган, чиқиши ночизикли боғланиш кўринишдаги жараён ҳолатини синфлаштиришнинг норавшан модели куйидагича тавсифланади.

Агар

$$(x_1^m = a_{11}^m \vee x_2^m = a_{12}^m \vee \dots \vee x_n^m = a_{1n}^m) \wedge$$

.....

$$\wedge (x_1^k = a_{11}^k \vee x_2^k = a_{12}^k \vee \dots \vee x_n^k = a_{1n}^k)$$

У ҳолда

$$y_m = b_{m0} + b_{m1} \frac{\sum_{j=1}^q \mu(x_1^{mj}) x_1^{mj}}{\sum_{j=1}^q \mu(x_1^{mj})} + b_{m2} \frac{\sum_{j=1}^q \mu(x_2^{mj}) x_2^{mj}}{\sum_{j=1}^q \mu(x_2^{mj})} + \dots + b_{m-1} \frac{\sum_{j=1}^q \mu(x_{n-1}^{mj}) x_{n-1}^{mj}}{\sum_{j=1}^q \mu(x_{n-1}^{mj})} + b_{mm} \frac{\sum_{j=1}^q \mu(x_n^{mj}) x_n^{mj}}{\sum_{j=1}^q \mu(x_n^{mj})}.$$

.Бунда  $m$ -ўқув танланмадаги объектлар сони,  $n$ -объектнинг белгилари сони.

Сушт шаклланган объектларни синфлаштириш норавшан моделини қуришда норавшан кластерлаш алгоритмидан фойдаланилди ва ушбу алгоритмнинг афзалликлари сифатида куйидагиларни кўрсатиш мумкин:

- Маълумотларнинг таркиби тўғрисидаги априори фаразларга (кластерлар бўйича эҳтимолий тақсимотнинг кўриниши ва параметрлари, зичлик марказлари, кластерлар сони) бўлган заруриятнинг йўқлиги.
- Кластерлар бўйича бўлиш натижаларининг тушунарли талқини: ўзаро бир-бирларига яқин бўлган элементлар кетма-кетлиги мавжуд бўлган ҳолларда элементлар битта кластерга киритилади.
- Кластерларнинг геометрик шакллари бўлган чекланишларнинг мавжуд эмаслиги.
- Алгоритмнинг бажарилиш вақти кирувчи векторлардаги компонентлар сонига кам даражада боғлиқлиги.

Ушбу алгоритм куйидаги бешта қадам асосида бажарилади.

**1-қадам.** *Ўқув танланмани аниқлаш:* Алгоритм ўқув танланмасини аниқлашдан бошланади:

$$\begin{array}{c|c} \begin{array}{c} x_1^1 \ x_2^1 \ \dots \ x_n^1 \\ x_1^2 \ x_2^2 \ \dots \ x_n^2 \\ \cdot \quad \cdot \quad \cdot \\ x_1^m \ x_2^m \ \dots \ x_n^m \end{array} & \begin{array}{c} | y_1 \\ | y_2 \\ | \dots \\ | y_m \end{array} \\ \hline & \cdot \end{array}$$

**2-қадам.** *Ўқув танланмани нормаллаштириш:* Кейинги қадамда ўқув танланмадаги объектлар параметрларининг ва улар чиқиш ҳолатларининг қийматлари устида минимакс нормаллаштириш амаллари бажарилади:

$$u_i^k = l \frac{x_i^k - x^{\min}}{x^{\max} - x^{\min}}; \quad v^k = l \frac{y^k - y^{\min}}{y^{\max} - y^{\min}};$$

бу ерда  $i = \overline{1, n}$ ,  $k = \overline{1, m}$ .

**3-қадам.** *Фазиффикациялаш:* Ушбу қадамда ўқув танланмасидаги нормаллаштирилган қийматлар устида фазиффикациялаш оператори бажарилади:

$$\mu^j(u_i^k) = \frac{1}{1 + \frac{u_i^k - c_j}{\sigma_j}}; \quad \mu^j(v^k) = \frac{1}{1 + \frac{v^k - c_j}{\sigma_j}}; \quad \mu^*(u_i^k) = \max_j \mu^j(u_i^k); \quad \mu^*(v^k) = \max_j \mu^j(v^k);$$

$$SP_k = \prod_{j=1}^n \mu^*(u_j^k) \cdot \mu^*(v^k).$$

бу ерда  $\sigma_j, c_j$  - параметрлар,  $j = \overline{1, l}$ .  $l$  – ораликни ифодаловчи сон.

**4-қadam.** *Норавшан қоидаларни танлаш:* Норавшан қоидаларни танлаш амали  $SP^k$  қийматларини нормаллаштириш ва нормаллаштирилган қийматлар устида фаззификация амалини бажариш орқали амалга оширилади:

4.1.  $SP^k$  ни нормаллаштириш:

$$\eta^k = l \frac{SP_k - SP^{\min}}{SP^{\max} - SP^{\min}}.$$

4.2. Фаззификация:

$$\mu^j(\eta^k) = \frac{1}{1 + \frac{\eta^k - c_j}{\sigma_j}}; \quad \mu^*(\eta^k) = \max_j \mu^j(\eta^k).$$

**5-қadam.** *Норавшан кластеризация усули ёрдамида қоидаларни кластерлаштириш:* Эквивалентлик муносабати асосида норавшан кластеризация усули ёрдамида  $\{SP_1, \dots, SP_n\}$  қоидалар тўпламини ўзаро кесишмайдиган эквивалентлик синфларига ажратилади. Яъни, икки элемент фақат ва фақат улар ўртасида жуфт ҳолда бир-бирларига яқин бўлган элементлар кетма-кетлиги мавжуд бўлган ҳоллардагина битта эквивалентлик синфига киритилади [4].

**6-қadam.** *Модел параметрларини созлаш:* Ушбу босқичда модел параметрларини нейрон тўрлар, генетик алгоритм ва эволюцион алгоритмлардан фойдаланган ҳолда созлаш – оптимал қийматларини аниқлаш мумкин. Бунда модел тегишлилик функциясининг параметрлари ҳамда  $w$  қоидалар вазнлари созланади.

### 3. Ҳисоблаш тажриба экспериментлари

Олинган натижалар асосида тажрибавий тадқиқотни амалга ошириш учун норавшан ёндашувли усулни қўллашда қуйидаги масалалар ечилди:

- Норавшан қоидалар базасини ҳосил қилиш ҳамда норавшан қоидалар базасининг норавшан модели параметрларини арилар колонияси алгоритми ёрдамида созлаш орқали қоидалар тўпламини қискартириш баробарида юқори самарадорликка (таниб олишнинг юқори фоиздаги кўрсаткичига) эришувчи норавшан моделни куриш.
- Турли масалаларда олинган натижаларнинг қиёсий таҳлилини жадвал кўринишда амалга ошириш.

1-жадвал

Норавшан модел,  $kNN$  ва  $SVM$  алгоритмлар натижалари солиштириш

	Норавшан	$kNN$	$SVM$
Ирис (Iris)	98.3	97.37	97.33

Ишлаб чиқилган алгоритмларни экспериментал синовдан ўтказиш мақсадида бир қатор алгоритмларни тестлаш учун кенг қўлланилувчи Iris dataset модел масаласини ечиш жараёни кўриб ўтилди.

Олинган натижаларни солиштирма таҳлилини ўтказиш мақсадида Ирис гулини синфлаштириш масаласини ечишда мавжуд  $kNN$  ( $K$ -NEAREST NEIGHBOR) ва  $SVM$  (Support Vector Machine) алгоритмларидан ҳам фойдаланилди. Қуйида (1-жадвал) норавшан алгоритм

асосида ва  $kNN$  ҳамда  $SVM$  синфлаштириш алгоритмлари асосида Ирис гулини синфлаштириш натижасида олинган натижалар келтириб ўтилган.

Синов жараёнлари ҳар бир масала учун 10 мартали *cross-validation* дан 10 мартадан текширув жараёнида фойдаланган ҳолда амалга оширилган [7]. Ҳисоблаш натижалари шуни кўрсатдики, кўриб ўтилган тест масалаларни ечишда норавшан моделнинг натижалари  $kNN$  ва  $SVM$  алгоритмлар асосида олинган натижалардан аниқлиги юқорироқ экан.

#### 4. Хулоса

Бугунги кунда норавшан мантиқий модел қуришнинг кўплаб усуллари кўриш мумкин. Улар бири-биридан қоидалар базасини қуриш усуллари, моделлар параметрларини оптималлаштириш усуллари, объектлар кириши ва уларнинг чиқишлари ўртасидаги боғлиқлик турларига кўра фарқланиши мумкин. Ушбу ишда норавшан мантиқий моделни норавшан кластерлаштириш усули ёрдамида қуриш масаласи кўриб ўтилди. Модел параметрларини оптималлаштириш масаласи жуда муҳим масалалардан ҳисобланиб, синфлаштириш сифатига сезиларли даражада ҳисса қўшади. Аммо, ушбу масала катта ҳажмли ҳисоблашлар талаб қилувчи масала ҳисобланади.

#### References:

- [1] Baskin II, Palyulin VA Zefirov NS Multilayer perceptrons in the study of dependence "structure-property" for organic compounds / Russian Chemical Journal (Journal of the Russian Chemical Society. Mendeleev). - 2006. - Т. 50. - P. 86-96.
- [2] Rotshteyn A.P. Intellectualnie texnologii identifikatsii: nechetskaya logika, geneticheskie algoritmi, neyronnie seti. -Vinnitsa: UNIVERSUM-Vinnitsa. 1999.-320 s.
- [3] Rutkovskaya D., Pilinskiy M., Rutkovskiy L. Neyronnie seti, geneticheskie algoritmi i nechetkie sistemi: Per.s polsk. I.D. Rudinskogo. -M.: Goryachaya liniya-Telekom, 2004. -452 s.
- [4] Mingliqulov Z.B. Algoritmi prinyatiya diagnosticheskix resheniy s ispolzovaniem neyronechetkix texnologiy// Uzb.jurn. "Problemy informatiki i energetiki". – Toshkent, 2011. - №1. – S. 71-76.
- [5] SHtovba S.D. "Vvedenie v teoriyu nechetskix mnojestv i nechetkuyu logiku". <http://www.matlab.exponenta.ru>.
- [6] Leonenkov A.V. Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH. – SPb., 2003.
- [7] Mingliqulov Z.B. Turli tegishlilik funktsiyalarida neyronoravshan to'rtinchi o'qitish va sinflashtirish masalalarini yechish // Materiali Respublikanskoy nauchno-texnicheskoy konferentsii «Sovremennoe sostoyanie i perspektivi razvitiya informatsionnix texnologiy». Tashkent. – 2011. –s. 347-352.
- [8] Abidin, T. and Perrizo, W. SMART-TV: A Fast and Scalable Nearest Neighbor Based Classifier for Data Mining. Proceedings of ACM SAC-06, Dijon, France, April 23-27, 2006. ACM Press, New York, NY, pp.536-540
- [9] Wang, H. and Bell, D. Extended k-Nearest Neighbours Based on Evidence Theory. The Computer Journal, Vol. 47 (6) Nov. 2004, pp. 662-672.
- [10] Yu, K. and Ji, L. Karyotyping of Comparative Genomic Hybridization Human Metaphases Using Kernel Nearest-Neighbor Algorithm, Cytometry 2002.

#### Web сайтлар

- [1]. [mnorinov@umail.uz](mailto:mnorinov@umail.uz)