

მსგავსების კომპიტური ზომები

თალიკო ჟვანია, მზია კიკნაძე, მანანა მალრაძე
საქართველოს ტექნიკური უნივერსიტეტი

რეზიუმე

ალგორითმული მსგავსების ზომების ერთ-ერთ ფართოდ გავრცელებულ კლასს წარმოადგენს ე.წ. კომპიტური მსგავსების ზომები. ნაშრომში განხილულია მსგავსების ალგორითმულ-კომპიტური ზომების ფორმირების ორი ალგორითმი, რომელიც ეფუძნება წინასწარ შერჩეულ ან ფორმირებულ მსგავსების ზომების სიმრავლეს, რომელთა საშუალებით შესაძლებელია სასწავლო ნაკრების რეალიზაციების პრაქტიკულად უშეცდომოდ ამოცნობა.

საკვანძო სიტყვები: ალგორითმი. კომპიტური მსგავსება.

1. შესავალი

ალგორითმული მსგავსების ზომების ერთ-ერთ ფართოდ გავრცელებულ კლასს წარმოადგენს ე.წ. კომპიტური მსგავსების ზომები, რომელიც ეყარება ხმის მიცემის მეთოდს, სადაც გამოყენებულია მსგავსების რამდენიმე (ერთზე მეტი) ზომა, რომლებიც იღებენ დამოუკიდებელ გადაწყვეტილებებს უცნობი რეალიზაციის ამა თუ იმ სახისადმი მიკუთვნების შესახებ. შემდგომში თითოეული მსგავსების ზომის მიერ მიღებული გადაწყვეტილება მიიჩნევა ერთ ხმად, ხოლო საბოლოო გადაწყვეტილების მისაღებად გამოიყენება სტანდარტული, სახეთა ამოცნობის თეორიაში ცნობილი, ხმათა უმრავლესობის პრინციპი, რომლის მიხედვითაც საბოლოო გადაწყვეტილება მიიღება იმის მიხედვით, თუ რომელი სახე მიიღებს მსგავსების ზომების ხმათა უმრავლესობას.

2. ძირითადი ნაწილი

ნაშრომში აღწერილია მსგავსების ალგორითმულ-კომპიტური ზომების ფორმირების ორი ალგორითმი.

ალგორითმი 1 - მსგავსების ალგორითმულ-კომპიტური ზომების ფორმირების პროცესი ეფუძნება წინასწარ შერჩეულ ან ფორმირებულ მსგავსების ზომების ρ სიმრავლეს, რომელთა საშუალებით უნდა მოხდეს სასწავლო ნაკრების რეალიზაციების პრაქტიკულად უშეცდომოდ ამოცნობა. სახეთა სასწავლო ნაკრები აღვნიშნოთ $\{X\}$ -ით, რომელიც შედგება ცალკეული სახეების სასწავლო ნაკრების ქვესიმრავლეებისაგან:

$$\{X\} = \{X_1\}, \{X_2\}, \dots, \{X_i\}, \dots, \{X_j\}.$$

მსგავსების ზომათა სიმრავლისთვის გვაქვს:

$$\{\rho\} = \rho_1, \rho_2, \dots, \rho_k, \dots, \rho_K; \quad K = \text{Card}\{\rho\}.$$

შედგენილია ცხრილი, რომელიც შეიძლება წარმოვადგინოთ $|\beta_{ki}^{mi}|$ მატრიცის სახით, სადაც β_{ki}^{mi} კოეფიციენტი ასახავს X_i^{mi} რეალიზაციის ρ_k მსგავსების ზომით ამოცნობის შედეგად მიღებულ სიტუაციებს, კერძოდ:

1. თუ X_i^{mi} რეალიზაცია ρ_k მსგავსების ზომის მიერ ამოცნობილია სწორად, მაშინ $\beta_{ki}^{mi} = 1$;
2. თუ X_i^{mi} რეალიზაცია ρ_k მსგავსების ზომის მიერ ამოცნობილ იქნა არასწორად, მაშინ $\beta_{ki}^{mi} = -1$;

სიმბოლოთი „ Δ “, აღვნიშნოთ პრედიკატი „მხოლოდ“. შემოტანილი აღნიშვნების გამოყენებით გადაწყვეტილების მიღების წესს, მაგალითად, მე-2 ნახაზის AC მონაკვეთისათვის, ექნება სახე:

$$X \in A_i, \text{ თუ } (\beta_{q_i} = 1 \wedge (\beta_{g_i} = 1 \cup \beta_{g_2})) \wedge (\forall \beta_{q_i} \neq 1 \wedge \forall \beta_{g_i} \neq 1), \quad (4)$$

სადაც β_{q_i} არის $\rho_{q_i} \in \{\rho_q\}$ მსგავსების ზომის, ხოლო $\beta_{g_i}, \beta_{g_2} - \rho_{g_i}, \rho_{g_2} \in \{\rho_g\}$ მსგავსების ზომების შესაბამისი კოეფიციენტები. ამასთან $\rho_{q_i}, \rho_{g_i}, \rho_{g_2}$ მსგავსების ზომები ზემოთ მიღებული წესის შესაბამისად ერთდროულად გადაფარავენ A_i სახის AC მონაკვეთზე მოთავსებულ რეალიზაციებს; β_{q_i} და β_{g_i} არის შესაბამისად $\rho_{q_i} \in \{\rho_q\}$ და $\rho_{g_i} \in \{\rho_g\}$ მსგავსების ზომების შესაბამისი კოეფიციენტები, რომლებიც არ გადაფარავენ განხილულ მონაკვეთს. A_i სახის ნებისმიერი სხვა მონაკვეთისათვის გადაწყვეტილების მიღების წესი ფორმირდება AC მონაკვეთის გადამფარავი მსგავსების ზომებისათვის შემუშავებული (4) გამოსახულების ანალოგიურად.

ზემოთ აღწერილი პროცედურები უნდა განვახორციელოთ $\{A\}$ სახეთა სიმრავლის ყველა ელემენტისათვის, რის შედეგადაც მიიღება ამ სიმრავლის ყოველი სახისათვის მსგავსების ზომათა კონკრეტული კომბინაციები.

თუ არ გამოიძებნა $\{\beta\}$ კოეფიციენტების შესაბამისი მსგავსების ზომების ρ სიმრავლიდან ისეთი ელემენტების გაერთიანება, რომლებიც ვერ უზრუნველყოფენ სახეთა A სიმრავლის თუნდაც ერთი ელემენტის სასწავლო (საკონტროლო) ნაკრების გადაფარვას, მაშინ აუცილებელია ρ სიმრავლის ელემენტების შეცვლა ან გაზრდა.

ალგორითმი 2: საწყის ინფორმაციას წარმოადგენს β კოეფიციენტების სიმრავლე. შემოვიტანოთ W_k კოეფიციენტები ყოველი ρ_k მსგავსების ზომისათვის

$$W_k = \frac{1}{I} \sum_i \frac{1}{M} \sum_{m_i} \beta_{ki}^{m_i}, \text{ სადაც } M_i = \text{Card}\{X_i\}; \quad i = \overline{1; I}. \quad (5)$$

განვიხილოთ შემდეგი სიტუაციები:

1) $W_v = 1$, რაც ნიშნავს, რომ ρ_v მსგავსების ზომა უშეცლოდ ამოიცნობს სახეთა A სიმრავლის ყველა რეალიზაციას. ცხადია, რომ ასეთ შემთხვევაში ρ_v მსგავსების ზომას ვირჩევთ ამოცნობის რეალური პროცესისათვის.

2) $W_k \neq 1, k = \overline{1; K}$. ამ შემთხვევისათვის განვსაზღვროთ W_k კოეფიციენტები ყოველი A_i სახის მიმართ ცალ-ცალკე. მაშინ (5) გამოსახულების ნაცვლად გვექნება:

$$W_{ki} = \frac{1}{M_i} \sum_{m_i} \beta_{ki}^{m_i}. \quad (6)$$

(6) გამოსახულებით განხორციელებული გამოთვლების მიხედვით მიიღება ცხრილი 1.

საწყისი ρ სიმრავლიდან მსგავსების ზომების შერჩევა (ρ^1 ქვესიმრავლის ელემენტების განსაზღვრა) ხდება შემდეგი თანმიმდევრობით: პირველად შერჩევა მსგავსების ზომა, მაგალითად ρ_{g_1} , რომელიც უშეცლოდ ამოიცნობს მოცემული სახეთა სიმრავლის ყველაზე მეტ ელემენტს.

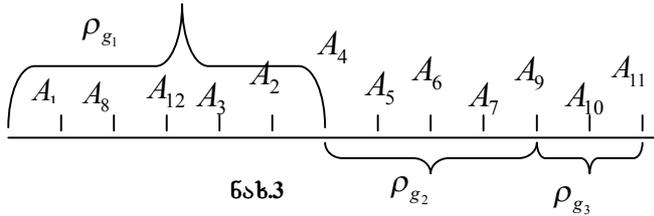
ცხრ.1

$\{A\} \setminus \{\rho\}$	ρ_1	ρ_2	ρ_3	...	ρ_k	...	ρ_K
A_1	W_{11}	W_{21}	W_{31}	...	W_{k1}	...	W_{K1}
A_2	W_{12}	W_{22}	W_{32}	...	W_{k2}	...	W_{K2}
...
A_i	W_{1i}	W_{2i}	W_{3i}	...	W_{ki}	...	W_{Ki}

აღვნიშნოთ ცხრილი 1-ის სვეტებში o_1 ინდექსით სახეთა სიმრავლის ის ელემენტები რომლის შესაბამისი $W_{ki} = 1$. ρ სიმრავლის მსგავსების ზომებიდან გამოვყოთ ρ_{g_1} მსგავსების ზომა, რომლისთვისაც სრულდება პირობა:

$$W_{g_1}^* = \max_k \left\{ \sum_{i=1} W_{k i} \right\}; \quad g_1 \leq K. \quad (7)$$

(7) პირობის მიხედვით პირველად აირჩევთ ის მსგავსების ზომა, რომელიც ყველაზე მეტი სახის რეალიზაციას ამოიცნობს სწორად. დავალაგოთ ღერძზე სწორად ამოცნობილი სახეები ერთად (ნახ.3). დავუშვათ, $A_1, A_8, A_{12}, A_3, A_2$ და A_4 სახეების რეალიზაციები სწორადაა ამოცნობილი ρ_{g_1} მსგავსების ზომის მიერ.



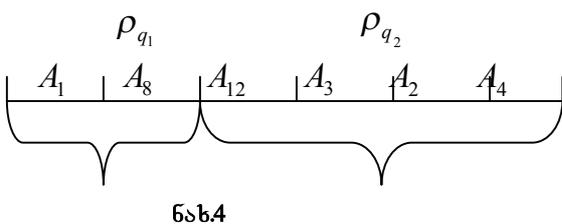
ρ_{g_1} მსგავსების ზომის მიერ დარჩენილი არასწორად ამოცნობილი სახეებისათვის სიმრავლიდან $\rho \setminus \rho_{g_1}$ განვსაზღვროთ მსგავსების ზომა, რომელიც აკმაყოფილებს (7) გამოსახულებას. აღვნიშნოთ ასეთი მსგავსების ზომა ρ_{g_2} -ით, მაშინ გვექნება:

$$W_{g_2}^* = \max_{k_2} \{ \sum_{i_2} W_{k_2 i_2} \}; \quad k_2 = \overline{1; K_2},$$

სადაც $K_2 = K - 1$; $k_2 \neq g_1$, k_2 ინდექსით აღნიშნულია მსგავსების ზომები, ρ_{g_1} მსგავსების ზომის გარდა, ხოლო i_2 ინდექსით თითოეულ სვეტში აღინიშნება ρ_{g_1} მსგავსების ზომის მიერ არასწორად ამოცნობილი, მაგალითად, $A_5, A_6, A_7, A_9, A_{10}, A_{11}$ სახეები, საიდანაც მიიღება ის სიმრავლე, რომლის შესაბამისი $W_{k_2 i_2} = 1$.

თუ ρ_{g_2} მსგავსების ზომის გამოყენების შემდეგ კიდევ დარჩა არასწორად ამოცნობილი სახეთა ქვესიმრავლე, ვირჩევთ (7) პირობის მიხედვით ისეთ მსგავსების ზომას, რომელიც გადაფარავს დარჩენილ, მაგალითად, A_{10} და A_{11} სახეებს. დავუშვათ, ასეთი მსგავსების ზომაა ρ_{g_3} . ეს პროცესი გავაგრძელოთ მანამ, სანამ დარჩენილ სახეთა ქვესიმრავლე ცარიელი გახდება.

დავუშვათ, რომ ზემოთ აღწერილი გადაფარვა ხორციელდება $\rho^1 \in \rho$ ქვესიმრავლის მსგავსების ზომებით. ცხადია, რომ ამოცნობის საიმპლობის ამაღლებისათვის სასურველი იქნება შეირჩეს მსგავსების ზომების კიდევ ერთი ქვესიმრავლე $\rho^2 \in \rho$, რომელიც მეორედ გადაფარავს სახეთა A სიმრავლის რეალიზაციებს, როგორც ეს ნაჩვენებია ნახ.4-ზე. აქ $A_1, A_8, A_{12}, A_3, A_2$ და A_4 სახეები მეორეაა გადაფარული ρ_{q_1} და ρ_{q_2} მსგავსების ზომებით, სადაც $\rho_{q_1}, \rho_{q_2} \in \{\rho^2\}$.



შევადგინოთ $\{\rho\}$ სიმრავლიდან მსგავსების ზომათა ისეთი ქვესიმრავლე, რომლის ყოველი ელემენტი შედგება მსგავსების ზომათა წყვილისაგან (შესაბამისად ρ^1 და ρ^2 ქვესიმრავლეებიდან), რომლებიც გადაფარავენ A სახეთა სიმრავლის ერთსა და იმავე სახეებს.

მსგავსების ზომათა კომიტეტის შერჩევა შემთხვევებისათვის, როდესაც $W_{ki} \neq 1 \Rightarrow W_k \neq 1$, მოითხოვს მსგავსების ზომათა შერჩევის განსხვავებულ პროცედურებს, რადგან სახეთა სიმრავლეში გვაქვს ისეთი სახეები, რომელთა სასწავლო ნაკრების ყველა რეალიზაციის სწორად ამოცნობა არ ხერხდება.

თითოეული იმ სახისათვის, რომლისთვისაც $W_{ki} \neq 1$, შემოვიტანოთ ρ_k მსგავსების ზომის ეფექტურობის მაჩვენებელი. დავუშვათ, რომ A_i სახის M_i რაოდენობის რეალიზაციებიდან სწორადაა ამოცნობილი Q_i რაოდენობის რეალიზაცია. აღვნიშნოთ p_{ki} -ით ამოცნობის ეფექტურობის მაჩვენებელი, რომელიც ტოლია:

$$p_{ki} = \frac{Q_i - G_i}{M_i}, \quad (8)$$

სადაც Q_i კოეფიციენტით წარმოდგენილია იმ შემთხვევათა რაოდენობა, როდესაც $\beta_{ki}^m = 1$, ხოლო G_i კოეფიციენტით წარმოდგენილია შემთხვევა, როდესაც $\beta_{ki}^m = 0$. როგორც წესი $Q_i > G_i$, ამიტომაც $p_{ki} \leq 1$.

შევადგინოთ ცხრილი, სადაც A სიმრავლის ყოველი A_i ელემენტისათვის და ρ სიმრავლის ყოველი ρ_k ელემენტისათვის ჩაწერილი გვექნება p_{ki} სიდიდის მნიშვნელობა (ცხრილი 2).

ცხრ2

$\{A_i\} \setminus \{\rho\}$	ρ_1	ρ_2	ρ_3	...	ρ_k	...	ρ_K
A_1	p_{11}	p_{21}	p_{31}	...	p_{k1}	...	p_{K1}
A_2	p_{12}	p_{22}	p_{32}	...	p_{k2}	...	p_{K2}
..
A_i	p_{1i}	p_{2i}	p_{3i}	...	p_{ki}	...	p_{Ki}

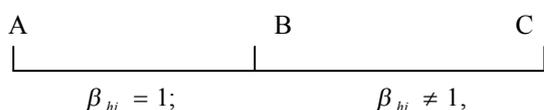
ცხადია, რომ p_{ki} სიმრავლის ელემენტები ფაქტობრივად წარმოადგენს სინშირულ პარამეტრებს.

მსგავსების ზომის შესარჩევად, მაგალითად A_i სახისათვის, განვიხილოთ შემდეგი პროცედურა (ალგორითმი).

მე-2 ცხრილის A_i რიგში განვსაზღვროთ ρ_h მსგავსების ზომა, რომელიც შესაბამისი p_{hi} -ისათვის დააკმაყოფილებს პირობას:

$$p_{hi}^* = \max_k \{p_{ki}\}; k = \overline{1; K}, h \leq K. \quad (9)$$

ა) დავყოთ A_i სახის სასწავლო-საკონტროლო ნაკრების რეალიზაციები ისე, რომ ერთ ნაწილში მოთავსდეს ის რეალიზაციები, რომლებსთვისაც $\beta_{hi} = 1$; მეორე ნაწილში კი - $\beta_{hi} \neq 1$ (ნახ. 5).



ნახ.5

A_i სახის რეალიზაციათა რაოდენობა M_i , მსგავსების ზომის ρ_k -ს მიხედვით, დაიყოფა ორ ნაწილად (ნახ.5).

$$M_i = M_{hi}^1 + M_{hi}^0, \quad (10)$$

სადაც M_{hi}^1 სიდიდეს შეესაბამება სწორად ამოცნობილი რეალიზაციები (მონაკვეთი AB), ხოლო M_{hi}^0 -ს - არასწორად ამოცნობილი (მონაკვეთი BC).

აღვნიშნოთ X_{hi}^1 -ით სწორად ამოცნობილი რეალიზაციათა სიმრავლე, ხოლო X_{hi}^0 -ით - არასწორად ამოცნობილი რეალიზაციათა სიმრავლე. მიღებული აღნიშვნების შესაბამისად გვექნება:

$$M_{hi}^1 = \text{Card}\{X_{hi}^1\}; M_{hi}^0 = \text{Card}\{X_{hi}^0\}.$$

ცხადია, რომ კმაყოფილდება პირობა: $\{X_{hi}^1\} \cap \{X_{hi}^0\} = \emptyset$.

ბ) (9) გამოსახულების გამოყენებით, X_{30}^0 სიმრავლის რეალიზაციებისათვის, შევარჩიოთ ρ_{h_1} მსგავსების ზომა ქვესიმრავლიდან $\{\rho\} \setminus \rho_h$:

$$p_{hi}^* = \max_{k_1} \{p_{k_1}^0\}, \quad k_1 = \overline{1; K_1}, \quad K_1 = K - 1,$$

სადაც k_1 ინდექსით აღინიშნება $\{\rho\} \setminus \rho_h$ მსგავსების ზომათა ქვესიმრავლის ელემენტები.

აქ შესაძლოა გვექნოდეს სამი სიტუაცია: პირველი, როდესაც ρ_{h_1} მსგავსების ზომა მთლიანად ფარავს X_{hi}^0 სიმრავლის რეალიზაციებს; მეორე, როდესაც ρ_{h_1} მსგავსების ზომა ნაწილობრივ ფარავს X_{hi}^0 სიმრავლის რეალიზაციებს; მესამე, როდესაც ρ_{h_1} მსგავსების ზომა ფარავს X_{hi}^0 სიმრავლის რეალიზაციებს მთლიანად და X_{hi}^1 რეალიზაციების ნაწილს (BC მონაკვეთს მთლიანად და AB მონაკვეთის ნაწილს).

სიტუაცია 1. $\beta_{hi} = 1$ მთელ BC მონაკვეთზე (ნახ.5). უცნობი X რეალიზაციის ამოცნობისათვის გვექნება:

$$X \in A_i \quad \text{თუ} \quad (\beta_{hi} = 1 \cap \beta_{hi} \neq 1) \cup (\beta_{hi} \neq 1 \cap \beta_{hi} = 1) \quad (11)$$

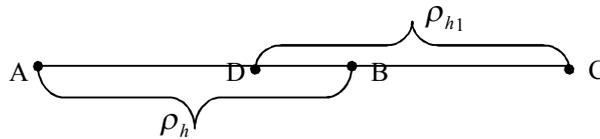
სიტუაცია 2. $\beta_{hi} \neq 1$ (სადაც β_{hi} - სიმბოლოთი აღნიშნულია პრედიკატი “მოიძებნება თუნდაც ერთი”). მიღებული აღნიშვნით ზემოთ მოცემული პირობა აღწერს სიტუაციას, როდესაც BC მონაკვეთზე მოიძებნება ერთი მაინც $X_i^{m_i}$ რეალიზაცია, რომლისთვისაც $\beta_{hi}^m \neq 1$.

ასეთ შემთხვევაში უნდა გავიმეოროთ ბ) პუნქტით მიღებული პროცედურები $\{\rho\} \setminus \rho_h$ სიმრავლისათვის.

სიტუაცია 3. თუ ρ_h მსგავსების ზომა BC მონაკვეთის გარდა ფარავს AC მონაკვეთის ნაწილსაც (ნახ.10), მაშინ გადაწყვეტილების მისაღებად შესაძლებელია გამოვიყენოთ შემდეგი წესი:

$$X \in A_i \quad \text{თუ} \quad (\beta_{hi} = 1 \cap \beta_{hi} \neq 1) \cup (\beta_{hi} = 1 \cap \beta_{hi} = 1) \cup (\beta_{hi} \neq 1 \cap \beta_{hi} = 1) \quad (12)$$

(12) გამოსახულების ჭეშმარიტება ცხადი ხდება ნახ.6-დან.



ნახ.6

აღსანიშნავია, რომ ტერმინი „გადაფარვა“ არ ნიშნავს (მსგავსების მიხედვით) ტერმინ „თანაკვეთას“, რაც თვალსაჩინოდ ჩანს ნახ.6-დან, სადაც ρ_h მსგავსების ზომის მიხედვით DB მონაკვეთი არსებობს მხოლოდ იმ შემთხვევაში, თუ AB მონაკვეთში მოთავსებული რეალიზაციების გადაადგილება მოხდება ρ_h მსგავსების ზომის მიხედვით, ე.ი. DB მონაკვეთში განვალაგებთ ამ მონაკვეთის იმ რეალიზაციებს, რომლებისთვისაც $\beta_{hi} = 1$.

სიტუაცია 2 და სიტუაცია 3-ით მოცემული პროცედურები უნდა განმეორდეს მანამ, სანამ BC მონაკვეთის სიგრძე გახდება ნულის ტოლი, რაც ნიშნავს, რომ აღარ იარსებებს A_i სახის რეალიზაციები, რომლებიც არასწორად იქნებიან ამოცნობილი.

A სიმრავლის ყველა სახისათვის ცალ-ცალკე ფორმირდება გადაწყვეტილების მიღების წესები, ანუ ყოველი სახისათვის უნდა მოხდეს ზემოთ აღწერილი პროცედურების შესრულება.

3. დასკვნა

ექსპერიმენტის შედეგებიდან გამომდინარე შეიძლება დავასკვნათ, რომ მსგავსების ზომების საჭირო რაოდენობა, რომელიც უზრუნველყოფს ერთი მსგავსების ზომის მიერ მიღწეულ შედეგზე უკეთეს შედეგს, დამოკიდებულია არასწორად ამოცნობილი რეალიზაციების რაოდენობაზე. ალგორითმი 1-ით და ალგორითმი 2-ით შერჩეული მსგავსების ზომების სიმრავლეები თითქმის ერთი და იმავე სიმძლავრის არიან, მაგრამ შეიძლება შედგებოდნენ სხვადასხვა ელემენტებისაგან. ამოცნობის საიმედოობის ზრდა მით უფრო მეტი და თვალსაჩინოა, რაც უფრო მცირეა საწყისი მსგავსების ზომით მიღებული ამოცნობის საიმედოობა. ამასთან, უნდა აღინიშნოს, რომ მკვეთრად კლებულობს ამოცნობის ოპერატიულობა.

4. ლიტერატურა

1. ვერულავა ო., ხუროძე რ., ამომცნობი სისტემების თეორიის საფუძვლები. სტუ, თბ., 2001.
2. Мазуров В.Д. Метод комитетов в задачах оптимизации и классификации. М., 1990.
3. Загоруйко Н. Г. Методы распознавания и их применения.. М., Советское Радио, 1972.

COMMITTEE SIMILARITY MEASURES

Zhvania Taliko, Kiknadze Mzia, Magradze Manana
Georgian Technical University

Summary

The most commonly used class of algorithmic similarity measures is so called the Committee Similarity Measures. There are considered two algorithms of formation of algorithmic and committee measures. These algorithms are based on preliminary chosen or already formed set of similarity measures. By means of these similarity measures are possible practically error-free recognition of realization of learning sets.

КОМИТЕТНЫЕ МЕРЫ СХОДСТВА

Жвания Т., Кикнадзе М., Маградзе М.
Грузинский Технический Университет

Резюме

Одним из широко распространенных классов алгоритмических мер сходства являются т.к. комитетные меры сходства. В статье рассмотрено формирование двух алгоритмов, алгоритмических и комитетных мер сходства. Для комитетных мер сходства были использованы существующие и сформированные меры сходства, с помощью которых возможно осуществление практически безошибочное распознавания реализации учебной выборки.