

შემავალი ინფორმაციის ფორმირება მანქანური დასწავლის ალგორითმებში

კრისტოფ პოე - მარსელის ცენტრალური უნივერსიტეტი,

ვალერიანე გელოვანი - საქართველოს ტექნიკური უნივერსიტეტი

რეზიუმე

განხილულია კლასიფიკაციის ტიპის ალგორითმებისათვის შემავალი ინფორმაციის ფორმირების საკითხები მანქანური დასწავლის სისტემებში. შემუშავებულია ლოგიკური სქემის მოდელი, რომლის დახმარებითაც შესაძლებელია შემავალი ინფორმაციის Y ვექტორსვეტის განსაზღვრა. მას დიდი მნიშვნელობა აქვს, რადგან მთლიანი ხის აგება გარკვეულწილად Y სვეტზეცაა დამოკიდებული. ნაშრომში წარმოდგენილია ბინარული მოდელი.

საკვანძო სიტყვები: ინტელექტუალური სისტემა. მანქანური დასწავლა. კლასიფიკაცია. ალგორითმი.

1. შესავალი

მანქანური დასწავლის ძირითადი მიზანია რაიმე სისტემაზე დაკვირვება. არსებული სისტემები შეიძლება იყოს მრავალნაირი: მათემატიკური, ბიოლოგიური, კომპიუტერული და ა.შ. მისი ერთ-ერთი სახეა ინტელექტუალური სისტემა. ეს სისტემები მოიცავს სხვადასხვა დარგებს. მანქანური დასწავლა ინტელექტუალური სისტემის ერთ-ერთი მნიშვნელოვანი ნაწილია, რაც თავის მხრივ საჭიროებს ისეთ ანალიზურ საკითხებს როგორცაა: შემავალი და გამომავალი ინფორმაცია, სისტემის მუშაობის ძირითადი პროცესები და პრინციპები. ასეთი ტიპის სისტემების მანქანურ დასწავლაში განსაზღვრულია მრავალი სხვადასხვა ტიპის ალგორითმი, რომელთაგან ერთ-ერთია *კლასიფიკაცია*. მნიშვნელოვანია ვიცოდეთ, როგორ და რა პრინციპით დავადგინოთ შემავალი ინფორმაცია ამ ტიპის ალგორითმებისათვის.

მანქანური დასწავლა ინტელექტუალური სისტემის ერთ-ერთი უმთავრესი კომპონენტია. მისი საშუალებით ხდება აღნიშნულ სისტემაზე გარკვეული დაკვირვებები. ასეთი ტიპის სისტემები გამოირჩევა გარკვეული სპეციფიკაციით, კერძოდ ის შეიძლება იყოს დახურული ან ღია ტიპის. დახურული ტიპის სისტემებში უცნობია ის პრინციპები და ალგორითმები, რომლითაც აღნიშნული სისტემა მუშაობს, შესაბამისად დაკვირვებები ხორციელდება მხოლოდ შემავალ და გამომავალ ინფორმაციაზე. მისგან განსხვავებით ღია ტიპის სისტემებში ცნობილია არა მხოლოდ შემავალი და გამომავალი ინფორმაცია, არამედ ის ალგორითმები და პრინციპები, რომელსაც იყენებს აღნიშნული სისტემა.

მანქანურ დასწავლაში გამოყენებული ალგორითმები დაჯგუფებულია გარკვეული სახით, როგორცაა კლასიფიკაციის, კლასტერიზაციის და ასოციაციის ალგორითმები. როგორც წესი, ასეთი ტიპის ალგორითმები ცალკე აღებული ვერ იძლევა ეფექტურ შედეგს, ამიტომ საჭირო ხდება მათი გარკვეული წესით შერწყმა. ამ ალგორითმების კომბინაციისათვის ერთ-ერთ მნიშვნელოვანი ფაქტორია შემავალი და გამომავალი ინფორმაცია. ერთი ალგორითმის გამომავალი ინფორმაცია შეიძლება იყოს შემავალი

ინფორმაცია მეორე ალგორითმისთვის. ამ შემთხვევაში მნიშვნელოვანია შემდეგი ძირითადი ფაქტორები:

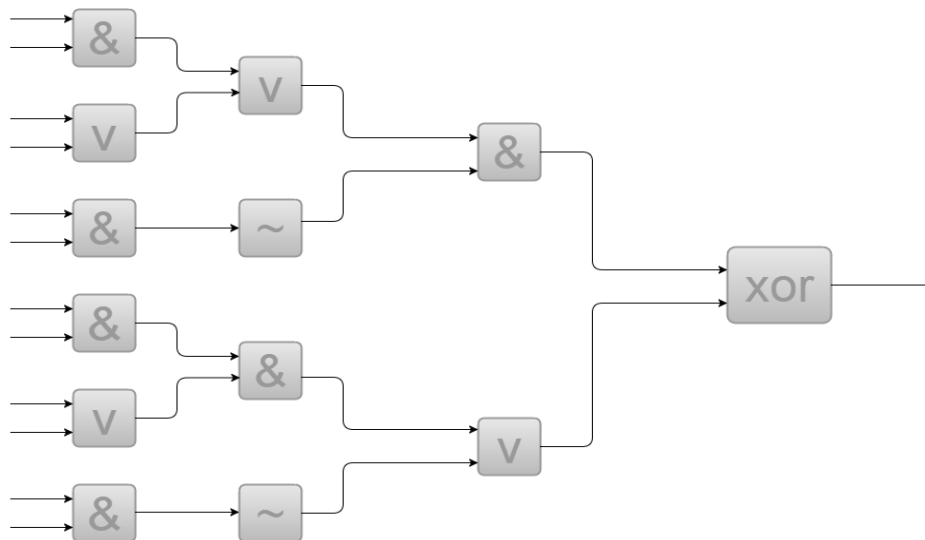
❖ *ალგორითმების სწორი შერჩევა* - მათი კომბინაციის შემთხვევაში უნდა განისაზღვროს თუ რომელი ალგორითმების კომბინაცია იქნება უკეთესი (კომბინაციაში შეიძლება იგულისხმებოდეს სხვადასხვა კატეგორიის ალგორითმების ერთმანეთში შერწყმაც);

❖ *ინფორმაციის წარმოდგენის წესი* - შესაძლებელი უნდა იყოს მოცემული ინფორმაციის წარმოდგენა როგორც კონკრეტული ალგორითმის შემავალი ინფორმაცია;

ინფორმაციის წარმოდგენის წესი შეიძლება გამოყენებულ იქნას არა მხოლოდ ისეთ სისტემებში სადაც მანქანურ დასწავლაში ჩადებულია ალგორითმების კომბინაცია, არამედ ცალსახად, ნებისმიერი ცალკე აღებული ალგორითმის შემთხვევაში.

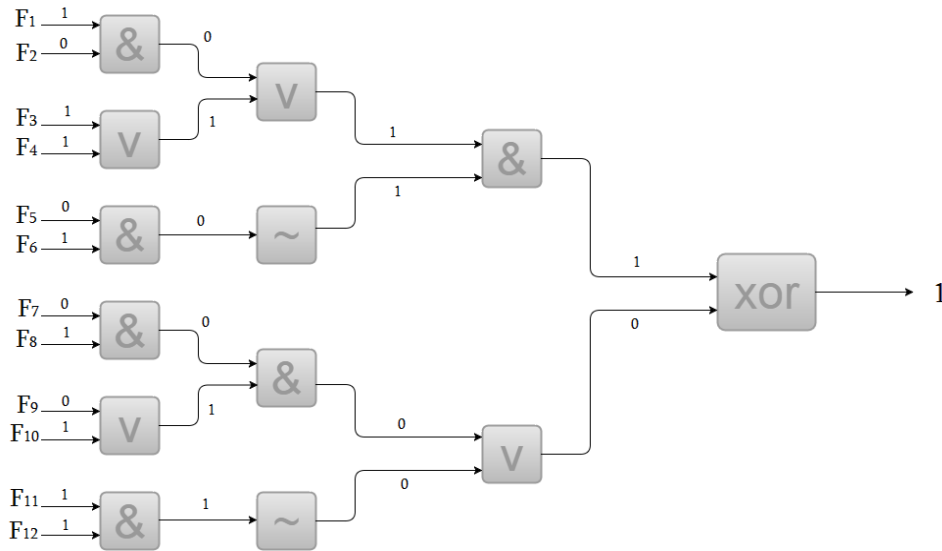
2. შემავალი ინფორმაციის ფორმირება

ნებისმიერი ალგორითმისათვის აუცილებელია შემავალი ინფორმაცია. როგორც უკვე აღვნიშნეთ, მიმდინარე ალგორითმი იყენებს მატრიცას, რომელიც შედგება კონკრეტული რაოდენობის ვექტორ-სტრიქონისა და Y ვექტორ-სვეტისაგან. ასევე წინასწარაა მოცემული დათვლილი საშუალო ენტროპიები. შემავალი ინფორმაციიდან ჩვენ შეიძლება ვიცოდეთ მხოლოდ მატრიცა, Y სვეტის გარდა. მისი ფორმირებისათვის ჩვენ გამოვიყენებთ ე.წ ლოგიკურ სქემას, რომელსაც შესასვლელზე მიეწოდება 0-ები და 1-ები, ხოლო გამოსასვლელზე იღებს 0-ს ან 1-ს, კონკრეტული ლოგიკური გამოსახულებიდან გამომდინარე. ასეთი სქემის შერჩევა დამოკიდებულია ოპერანდების რაოდენობაზე, მაგალითისთვის შეგვიძლია მოვიყვანოთ ჩვენ მიერ შემუშავებული ლოგიკური სქემა, რომელიც განსაზღვრულია 12 ოპერანდზე:



ნახ.1. ლოგიკური სქემა

მოცემულ სქემაში გამოყენებულია 4 ტიპის ატომური ოპერატორი: „და“ (&) , „ან“ (V), „უარყოფა“ (~), „ბიტური ან“ (XOR). შესასვლელზე მიეწოდება 0-ები და 1-ები, საბოლოოდ კი მივიღებთ 0-ს ან 1-ს ამ სქემიდან გამომდინარე. შეგვიძლია მოვიყვანოთ კონკრეტული მაგალითი:



ნახ.2. ლოგიკური სქემის გამოყენების კონკრეტული მაგალითი

3. კლასიფიკაციის ერთ-ერთი ალგორითმი

როგორც უკვე აღვნიშნეთ, კლასიფიკაცია აერთიანებს კონკრეტული ტიპის ალგორითმებს. მისი ერთ-ერთი კლასიკური ალგორითმია გადაწყვეტილებათა ხე (Decision Tree) [1]. ჩვენ განვიხილავთ ისეთ გადაწყვეტილებათა ხეს, სადაც მონაცემები წარმოდგენილია ორობითი სახით. ალგორითმისათვის წინასწარაა ცნობილი/მოცემული:

- მატრიცა (რომელიც აგებულია გარკვეული ვექტორ-სვეტების საფუძველზე);
- საწყისი ენტროპია (გამოთვლილი სახით);

ალგორითმი მუშაობს შემდეგი პრინციპით:

- თითოეული ვექტორ-სვეტისათვის გამოვთვალოთ საშუალო ენტროპია AE_i ;
- გამოთვლილი საშუალო ენტროპიებიდან ამოვარჩიოთ მინიმალური, რის შედეგადაც მოხდება წინა ეტაპზე მიღებული მატრიცის გაყოფა მარცხენა და მარჯვენა ქვე მატრიცებად, ხოლო არჩეული ვექტორ სვეტი კი იქნება ხის მიმდინარე კვანძი;
- გავიმეოროთ ეს პროცესი მანამ, სანამ არ მოხდება მთელი ხის აგება;
- ახლა განვსაზღვროთ თუ როგორ ხდება საშუალო ენტროპიის დათვლა თითოეული ვექტორ-სვეტისათვის:

1) f_i ვექტორ-სვეტისათვის დავთვალოთ დადებითი და უარყოფითი ელემენტების რაოდენობა მარცხენა და მარჯვენა კვანძისათვის (თითოეული f_i წარმოადგენს ხის კვანძს, რომელსაც გააჩნია მარცხენა (დადებითი კვანძი) და მარჯვენა (უარყოფითი კვანძი) კვანძები). მარცხენა კვანძისთვის დადებითი ელემენტების რაოდენობა ტოლია იმ 0 ების რაოდენობის, რომლის შესაბამისი Y არის 0. მარცხენა კვანძისთვის უარყოფითი ელემენტების რაოდენობა ტოლია იმ 0 ების რაოდენობის, რომლის შესაბამისი Y არის 1. მარჯვენა კვანძისთვის დადებითი ელემენტების რაოდენობა ტოლია იმ 1 ის ტოლი მნიშვნელობების რაოდენობის, რომლისთვისაც შესაბამისი Y არის 0. მარჯვენა

კვანძისთვის უარყოფითი ელემენტების რაოდენობა ტოლია იმ 1 ის ტოლი მნიშვნელობების რაოდენობის, რომლისთვისაც შესაბამისი Y არის 1;

2) ენტროპია მოცემულია შემდეგი ფორმულით (რადგან საუბარია ორობით სისტემაზე, აუცილებელია ორივე ენტროპიის გამოთვლა):

$$I_i(p_0, n_0) = -\frac{p_0}{p_0+n_0} \log_2\left(\frac{p_0}{p_0+n_0}\right) - \left(1 - \frac{p_0}{p_0+n_0}\right) \log_2\left(1 - \frac{p_0}{p_0+n_0}\right) \text{ და}$$

$$I_i(p_1, n_1) = -\frac{p_1}{p_1+n_1} \log_2\left(\frac{p_1}{p_1+n_1}\right) - \left(1 - \frac{p_1}{p_1+n_1}\right) \log_2\left(1 - \frac{p_1}{p_1+n_1}\right);$$

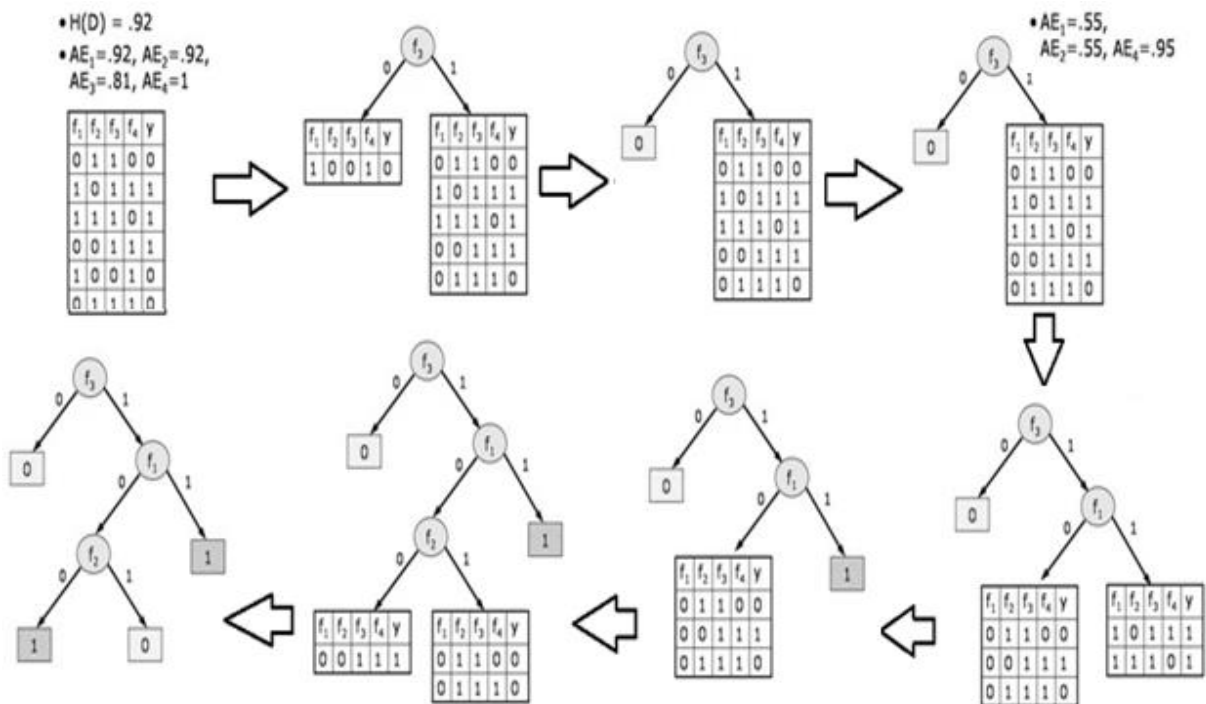
3) ბოლოს უნდა გამოვთვალოთ AE_i შემდეგნაირად:

$$AE_i = \frac{p_0+n_0}{p+n} I_i(p_0, n_0) + \frac{p_1+n_1}{p+n} I_i(p_1, n_1);$$

ამ 3 ეტაპის შემდეგ ვარჩევთ ისეთ AE_i რომელიც არის მინიმალური, ვფოფთ მატრიცას ორ ნაწილად (მარცხენა და მარჯვენა ქვე მატრიცებად. გაყოფა ხდება არჩეული ვექტორ სვეტის საფუძველზე) და მთელ პროცესს ვიმეორებთ მანამ, სანამ არ მოხდება მთელი ხის აგება [2].

ნახ.3. გადაწყვეტილებათა ხის კონკრეტული მაგალითი

4. დასკვნა



კვლევის საფუძველზე შემუშავებულ იქნა ლოგიკური სქემის მოდელი, რომლის დახმარებითაც შესაძლებელია შემავალი ინფორმაციის Y ვექტორსვეტის განსაზღვრა. მას დიდი მნიშვნელობა აქვს, რადგან მთლიანი ხის აგება გარკვეულწილად Y სვეტზეცაა დამოკიდებული. რა თქმა უნდა, ჩვენ განვიხილეთ ბინარული მოდელი. არაბინარული მოდელის შემთხვევაში შეგვიძლია მთლიანი ინფორმაცია დავიყვანოთ ბინარულზე ან პირდაპირ ავაგოთ გადაწყვეტილებათა ხე. არაბინარული მოდელის შემუშავება ცალკე განხილვის თემაა.

ლიტერატურა - References – Литература:

1. Artificial Intelligence. Copyright © 2004 by Massachusetts Institute of Technology.
2. ზოსიკაშვილი ზ., გელოვანი ვ. (2018). კიბერუსაფრთხოების სისტემების აგების პრინციპების შემუშავება და კვლევა პროგრამული უზრუნველყოფის სუსტი ადგილების ანალიზისა და მოდელირებისათვის. სტუ-ს შრ.კრ. „მას" N1(25). თბილისი. გვ.21-28.

**INPUT INFORMATION GENERATION IN MACHINE
LEARNING ALGORITHMS**

Pouet Christophe – Central school of Marseille,
Gelovani Valeriane - Georgian Technical University

Summary

Questions of formation of the input information for algorithms of classification type in machine learning systems are considered. A model of a logical scheme has been developed, with the help of which it is possible to determine the Y vector-column of the input information. This is of great importance, since the construction of the entire decision tree depends on the Y column. A binary model is presented in this paper.

**ФОРМИРОВАНИЕ ВХОДНОЙ ИНФОРМАЦИИ В АЛГОРИТМАХ МАШИННОГО
ОБУЧЕНИЯ**

Пое К. – Центральная школа Марселя,
Геловани В. - Грузинский Технический Университет

Резюме

Рассматриваются вопросы формирования входной информации для алгоритмов типа классификации в системах машинного обучения. Разработана модель логической схемы, с помощью которой можно определить Y вектор-столбец входной информации. Это имеет большое значение, так как построение всего дерева принятия решений зависит от столбца Y. В работе представлена двоичная модель.