

მანქანური სწავლების კლასიფიკაციის ალგორითმებში შემავალი ინფორმაციის ფორმირება

ზურაბ ბოსიკაშვილი, დავით ჭოხონელიძე
საქართველოს ტექნიკური უნივერსიტეტი

რეზიუმე

მანქანური სწავლების ძირითადი მიზანია რაიმე სისტემაზე დაკვირვება. არსებული სისტემები შეიძლება იყოს მრავალნაირი: მათემატიკური, ბიოლოგიური, კომპიუტერული და ა.შ. მისი ერთ-ერთი სახეა ინტელექტუალური სისტემა. ეს სისტემები მოიცავენ სხვადასხვა დარგებს. მანქანური სწავლება ინტელექტუალური სისტემის ერთ-ერთი მნიშვნელოვანი ნაწილია, რაც თავისმხრივ საჭიროებს ისეთ ანალიზურ საკითხებს როგორცაა: შემავალი და გამომავალი ინფორმაცია, სისტემის მუშაობის ძირითადი პროცესები და პრინციპები. ასეთი ტიპის სისტემების მანქანურ სწავლებაში განსაზღვრულია უამრავი სხვადასხვა ტიპის ალგორითმი, რომელთაგან ერთ-ერთია კლასიფიკაცია. მნიშვნელოვანია ვიცოდეთ, როგორ და რა პრინციპით დავადგინოთ შემავალი ინფორმაცია ამ ტიპის ალგორითმებისათვის. მიმდინარე სტატია განიხილავს კლასიფიკაციის ტიპის ალგორითმებისათვის შემავალი ინფორმაციის ფორმირებას.

საკვანძო სიტყვები: ინტელექტუალური სისტემა. მანქანური სწავლება. კლასიფიკაციის ალგორითმები.

1. შესავალი

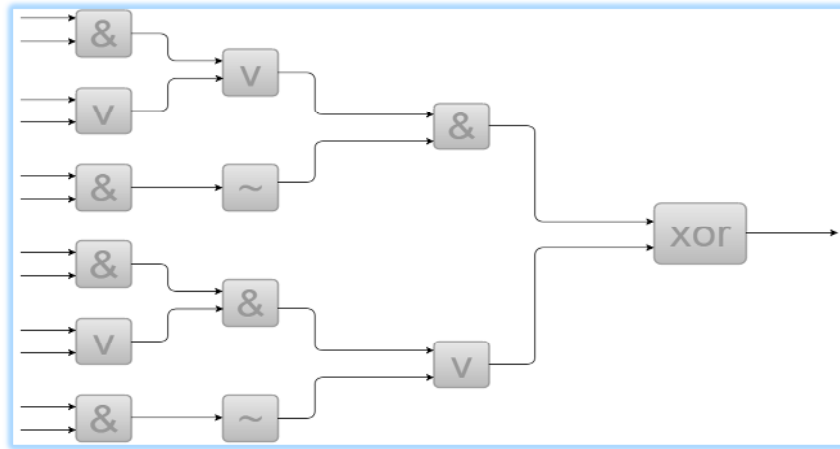
მანქანური სწავლება წარმოადგენს ინტელექტუალური სისტემის ერთ-ერთ უმთავრეს კომპონენტს. მისი საშუალებით ხდება აღნიშნულ სისტემაზე გარკვეული დაკვირვებები. ასეთი ტიპის სისტემები გამოირჩევიან გარკვეული სპეციფიკაციით, კერძოდ ის შეიძლება იყოს დახურული ან ღია ტიპის. დახურული ტიპის სისტემებში უცნობია ის პრინციპები 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 სწავლებაში გამოყენებული ალგორითმები დავაგუფებულია გარკვეული სახით. ზოგიერთი მათგანია: 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 მნიშვნელოვან ფაქტორს წარმოადგენს შემავალი და გამომავალი ინფორმაცია. ერთი ალგორითმის გამომავალი ინფორმაცია შეიძლება წარმოადგენდეს შემავალ ინფორმაციას მეორე ალგორითმისთვის. ამ შემთხვევაში მნიშვნელოვანია შემდეგი ძირითადი ფაქტორები:

- ალგორითმების სწორი შერჩევა - მათი კომბინაციის შემთხვევაში უნდა განისაზღვროს 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 სწავლებაში ჩადებულია ალგორითმების კომბინაცია, არამედ ცალსახად, ნებისმიერი ცალკე აღებული ალგორითმის შემთხვევაში.

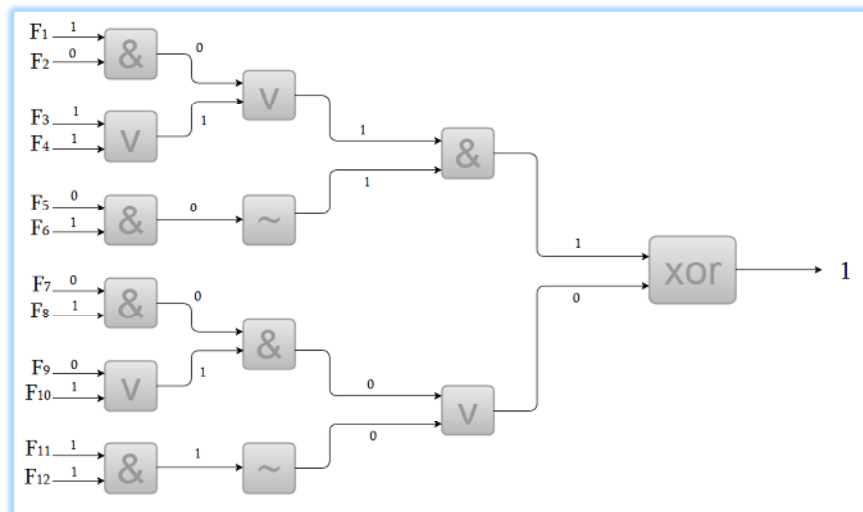
2. შემავალი ინფორმაციის ფორმირება

ნებისმიერი ალგორითმისათვის აუცილებელია შემავალი ინფორმაცია. როგორც უკვე აღვნიშნეთ მიმდინარე ალგორითმი იყენებს მატრიცას, რომელიც შედგება კონკრეტული რაოდენობის ვექტორ-სვეტისაგან 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 ჩვენს მიერ შემუშავებული ლოგიკური სქემა, რომელიც განსაზღვრულია 12 ოპერანდზე (ნახ.1).



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).



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 ხეები (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 ქვემატრიცებად, 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):

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

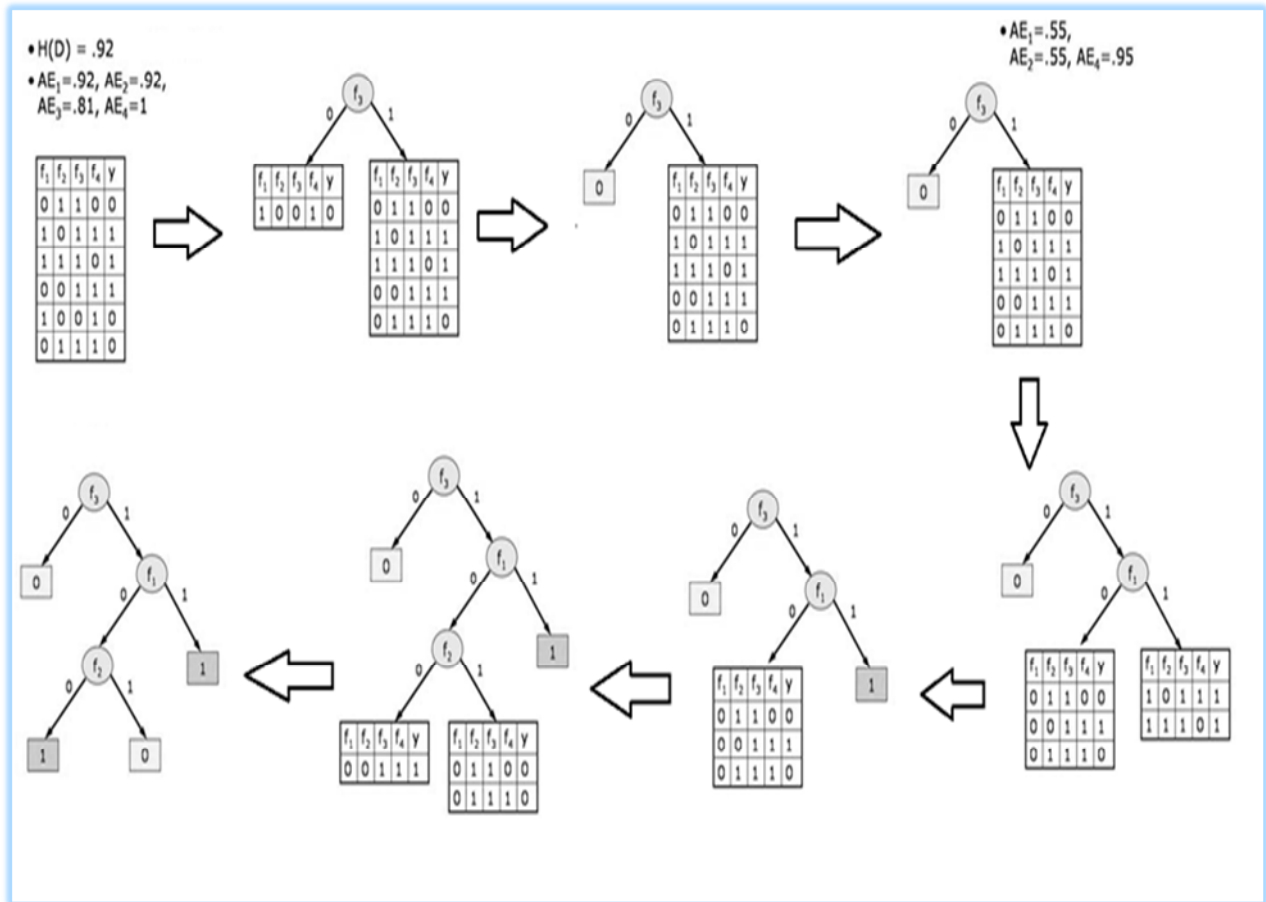
და

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

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, ვფოთო 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.



ნახ.3. გადაწყვეტილებათა ხე (კონკრეტული მაგალითი)

4. დასკვნა

კვლევის საფუძველზე შემუშავებულ იქნა ლოგიკური სქემის მოდელი, რომლის დახმარებითაც შეგვიძლია განვსაზღვროთ შემაჯალი ინფორმაციის y ვექტორ-სვეტი. მის განსაზღვრას დიდი მნიშვნელობა აქვს, რადგან მთლიანი ხის აგება გარკვეულწილად y სვეტზეცაა დამოკიდებული. რა თქმა უნდა ჩვენ განვიხილეთ ბინარული მოდელი, არა-ბინარული მოდელის შემთხვევაში შეგვიძლია მთლიანი ინფორმაცია დავიყვანოთ ბინარულზე ან პირდაპირ ავაგოთ გადაწყვეტილებათა ხე. არა-ბინარული მოდელის შემუშავება არ წარმოადგენს მიმდინარე სტატიის კვლევის საგანს, აქედან გამომდინარე ის შეიძლება განხილულ იქნას ცალკე.

ლიტერატურა:

1. Artificial Intelligence. (2004). Copyright © 2004 by Massachusetts Institute of Technology.

**INPUT INFORMATION GENERATION IN MACHINE LEARNING
CLASIFICATION ALGORITHMS**

Bosikashvili Zurab, Chokhonelidze David

Georgian Technical University

Summary

One of the main purpose of machine learning is observating the system. There exists many kinds of system: Mathmetical, Biological, Informational system and etc. One kind of such system is intelligence system. These systems are used in many industries. Machine learning is one of the main part of intelligence system which includes such questions: Input and output information, main processes of system. Such systems' machine learning defines many kind of algorithms. Such systems' machine learning defines many kind of algorithms. One kind of algorithm of this system is clasification. It's important to know how to generate input information for that kind of algorithms. Current article discusses how to generate input information of clasification algorithms.

**ФОРМИРОВАНИЕ ВХОДЯЩЕЙ ИНФОРМАЦИИ В АЛГОРИТМАХ
КЛАССИФИКАЦИИ МАШИННОГО ОБУЧЕНИЯ**

Босикашвили З., Чохонелидзе Д.

Грузинский Технический Университет

Резюме

Основная цель машинного обучения- наблюдение за какой- нибудь системой. Эти системы могут быть разнообразными: математические, биологические, компьютерные и т.д. Один из видов системы-интеллектуальная система. Эти системы используются в разных отраслях. Машинное обучение является важной частью интеллектуальной системы, которая включает в себя такие вопросы анализа, как входящая и выходящая информация, основные принципы и процессы работы системы. В машинном обучении для такого типа систем определены многие типы алгоритмов, один из которых- классификация. Важно знать как и каким принципом установить входящую информацию для такого типа алгоритмов. В представленной статье рассматривается формирование входящей информации для алгоритмов типа классификации.