**Klasyfikator**

Dokumentacja projektu

**Grupa projektowa:**  
Kamil Żak

Paweł Mitruś

Łukasz Napora

Mateusz Bukowski

Spis treści

[1. Cel dokumentu 3](#_Toc408757325)

[2. Środowisko programowania 3](#_Toc408757326)

[3. Opis problemu 3](#_Toc408757327)

[3.1. Reprezentacja i sposób klasyfikowania symboli 3](#_Toc408757328)

[3.2. Konstrukcja i symulacja automatu 4](#_Toc408757329)

[3.3. Optymalizacja macierzy przejścia automatu 5](#_Toc408757330)

[4. Etap I 5](#_Toc408757331)

[4.1. Generowanie zbioru uczącego 5](#_Toc408757332)

[4.2. Tworzenie funkcji przejścia 5](#_Toc408757333)

[4.3. Symulacja działania automatu 5](#_Toc408757334)

[4.4. Funkcja błędu 5](#_Toc408757335)

[5. Etap II 6](#_Toc408757336)

[5.1. Generowanie zbioru uczącego 6](#_Toc408757337)

[5.2. Tworzenie funkcji przejścia 6](#_Toc408757338)

[5.3. Symulacja działania automatu 6](#_Toc408757339)

[5.4. Funkcja błędu 6](#_Toc408757340)

[6. Etap III 6](#_Toc408757341)

[6.1. Generowanie zbioru uczącego 6](#_Toc408757342)

[6.2. Tworzenie funkcji przejścia 6](#_Toc408757343)

[6.3. Symulacja działania automatu 7](#_Toc408757344)

[6.4. Funkcja błędu 7](#_Toc408757345)

[7. Format uruchomieniowy programu 7](#_Toc408757346)

# Cel dokumentu

Celem dokumentu jest przedstawienie użytkownikowi sposobu działania Klasyfikatora. Dokument skupia się na wyjaśnieniu kwestii teoretycznej, metody rozwiązywania poszczególnych problemów oraz implementacji.

# Środowisko programowania

Klasyfikator został zaimplementowany w środowisku MatLab – głównym powodem wyboru tego środowiska, były liczne operacje na macierzach. W ten sposób, z wiedzą że MatLab dysponuje optymalnymi algorytmami do wyżej wspomnianych działań, staraliśmy się podnieść wydajność programu.

# Słownik pojęć

|  |  |
| --- | --- |
| Pojęcie | Znaczenie |
| Klasa |  |
| Symbol |  |
| Reprezentant |  |
| Cecha | Pewna wielkość matematyczna opisująca jedną z właściwości charakterystycznych dla symbolu z danej klasy |
| Liczba podziałów przedziału [0, 1] | Liczba, która informuje na ile podprzedziałów musimy dzielić przedział [0, 1] |
| Etap | Poszczególny etap oddawania projektu. Każdy poszczególny etap odnosi się do dwóch różnych zaimplementowanych automatów (z symbolami obcymi oraz bez symboli obcych) |
|  |  |

# Opis problemu

Problemem jest implementacja automatu klasyfikującego symbole z zadanych zbiorów do odpowiadających im klas na podstawie przedstawionych cech symboli.

Poszczególne etapy: 1, 2 oraz 3 różnią się wzajemnie typem implementowanego automatu. Odpowiednio dla:

1. Automat deterministyczny
   1. Bez symboli obcych
   2. Z symbolami obcymi
2. Automat niedeterministyczny
   1. Bez symboli obcych
   2. Z symbolami obcymi
3. Automat rozmyty
   1. Bez symboli obcych
   2. Z symbolami obcymi

## Reprezentacja i sposób klasyfikowania symboli

Do automatu wraz z każdym zbiorem wprowadzane są następujące wartości:

* Klasy: Zadany jest skończony zbiór klas , do których finalnie przypisywane są testowane symbole.
* Symbole: To reprezentanci każdej z klas. Zadany jest skończony zbiór symboli   
  . Po próbach rozpoznania symbolu są one przypisywane do odpowiednich klas.
* Cechy: To pewne określone właściwości każdego z symboli. Na ich podstawie decydujemy, czy dany symbol klasyfikuje się do danej klasy. Zadany jest skończony zbiór cech

Cechy

Symbole

Klasa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Symbol | C1 | C2 | … | Cn |
| A1.1 | 2 | 14 | 5,1 | 22 |
| A1.2 | 0 | 12 | 4,8 | 30 |
| … | 1,9 | 9 | 7 | 25 |
| A1.n | 0,5 | 11 | 6 | 26 |
| … |  |  |  |  |
| An.1 |  |  |  |  |
| ... |  |  |  |  |
| An.n |  |  |  |  |

## Konstrukcja i symulacja automatu

Każdy z etapów możemy opisać poniższym, ogólnym algorytmem. Kwestie, pomiędzy którymi poszczególne etapy się różnią, a w zasadzie są kluczowymi, zostaną opisane w punktach 4, 5, 6.

* + 1. Generowanie/wczytywanie zbioru uczącego – każdy ze zbiorów zawiera wartości dla symboli oraz cech.
    2. Generowanie/wczytywanie zbioru treningowego – są rozmiaru 1/3 ilości symboli odpowiadających im zbiorów uczących.
    3. Generowanie automatu – różne w zależności od etapu, opisane w punktach 4, 5, 6.
    4. Symulacja przejścia automatu dla zbioru uczącego– różne w zależności od etapu, opisane w punktach 4, 5, 6.
    5. Optymalizacja macierzy przejścia automatu – wykonywana jest za pomocą algorytmu PSO bazując na zbiorze uczącym. Dokładny odpis w punkcie 3.3.
    6. Obliczanie błędu – różne w zależności od etapu, opisane w punktach 4, 5, 6.

## Optymalizacja macierzy przejścia automatu

Optymalizacji dokonujemy algorytmem PSO (Particle Swarm Optimization, Optymalizacja Rojem Cząstek), który uprzednio zaimplementowaliśmy.

OPIS PSO

# Etap I

Różnice w automatach pomiędzy tymi bez symboli obcych, a automatami przyjmującymi zbiory z symbolami obcymi, zostaną opisane jako dodatki do poszczególnych punktów w celu uniknięcia kopiowania dużych ilości tekstu.

## Generowanie\wczytywanie zbioru uczącego

* + 1. Dla każdej z klas jest generowany losowo dokładnie jeden wektor cech rozkładem jednostajnym. Przedział losowania wartości ograniczony jest z góry i z dołu przez wczytywane parametry.
    2. Wylosowany wektor wartości cech dla klasy jest powielany w zależności od wczytanej liczby symboli.
    3. Następnie wartości w zbiorze zaburzane są szumem pochodzącym z rozkładu normalnego. Szum generowany jest losowo, a wartość oczekiwana oraz odchylenie standardowe są parametrami wejściowymi dla funkcji generującej.
    4. W kolejnym kroku zbiór cech jest normalizowany do przedziału [0,1]
    5. W zależności od wczytanej liczby podziałów przedziału [0, 1], dla każdej wartości cechy w zbiorze, przypisywana jest całkowita liczba porządkowa. Nowa wartość jest równa numerowi przedziału, w którym zawierała się poprzednia wartość cechy.

Dla przykładu: jeżeli liczba podziałów przedziału [0, 1] wynosi 5, otrzymujemy zbiór przedziałów

Jeżeli cecha to zostaje jej przypisana wartość ‘*i’*.

* + 1. Następnie, tworzona jest kolumna symboli w zależności od wczytanych parametrów (liczba klas, liczba symboli w klasie). Klasy zawsze przedstawiane są jako 1, 2, …, n.

**Automat z elementami obcymi:**

W przeciwieństwie do punktów 1 i 2 wektory cech dla symboli obcych nie są powielane – dla każdego symbolu losowany jest osobno wektor cech.

**Wczytywanie zbioru z pliku:**

Wykonujemy tylko kroki opisane w punktach 4 i 5.

## Generowanie/wczytywanie zbioru treningowego

* + 1. Dla każdej z klasy wybieramy losowo 1/3 wektorów cech (symboli z tej klasy) ze zbioru uczącego.
    2. Łączymy wylosowane wektory cech w nową macierz zbioru treningowego.

## Tworzenie funkcji przejścia

Automat jest deterministyczny, więc losowany jest tylko jeden stan akceptujący dla każdego symbolu.

## Generowanie automatu

## Symulacja przejścia automatu dla zbioru uczącego

## Obliczanie błędu

* + 1. Otrzymujemy wektor stanów po symulacji działania automatu
    2. Jeżeli symbol nie został rozpoznany, nie zwiększamy licznika rozpoznanych symboli, a w przypadku gdy został rozpoznany zwiększamy go o 1
    3. Obliczenia powtarzane są dla każdego symbolu
    4. Liczony jest procentowy błąd w klasyfikowaniu symboli

# Etap II

## Generowanie zbioru uczącego

* + 1. Zbiór cech losowany rozkładem jednostajnym
    2. Zbiór wylosowanych wartości zaburzany rozkładem Gaussa
    3. Zbiór cech jest normalizowany do przedziału [0,1]
    4. W zależności od ilości podanych przedziałów, dzielimy na tyle zadany przedział [0,1]
    5. Przypisujemy do każdej z cech wartość (porządkowa liczba całkowita) w zależności od tego w którym przedziale się znajduje.

## Tworzenie funkcji przejścia

Z racji że mamy do czynienia z automatem niedeterministycznym, losujemy kilka (od 0 do 3) stanów akceptujących dla każdego symbolu.

## Symulacja działania automatu

## Funkcja błędu

# Etap III

## Generowanie zbioru uczącego

* + 1. Zbiór cech losowany rozkładem Gaussa
    2. Zbiór cech jest normalizowany do przedziału [0,1]
    3. Każda z cech będzie rozszerzana do wektora o długości równej ilości przedziałów
    4. W wektorze umieszczamy wartości wg wzoru , gdzie to wartość cechy ze zbioru postaci w punkcie 2 oraz

Zasadnicza różnica polega na tym, że w wypadku słów rozmytych cecha nie jest reprezentowana przez jedną liczbę, a przez wektor. Zmienia to strukturę zbioru cech, z 2 wymiarowego na 3 wymiarowy.

## Tworzenie funkcji przejścia

Z racji że mamy do czynienia z automatem rozmytym losowane są wartości z przedziału [0 liczba.podzialow].

## Symulacja działania automatu

* + 1. Do automatu zostaje przekazana macierz cech danego symbolu.
    2. Zaczynamy od stanu początkowego stan.poczatkowy = [1 0 0 ... 0 ]
    3. Dla każdej z cech iterujemy po wektorze ją reprezentującym
    4. Przy użyciu funkcji:

Iterujemy przez wszystkie wartości w stan.poczatkowy i następnie z użyciem funkcji

tworzymy stan wejściowy dla kolejnej funkcji przejścia automatu.

* + 1. Zwracamy stworzony wektor stanów

W poprzednich etapach do symulacji automatu przekazywane były reprezentacje symbolu (wektor cech dla symbolu). W tym przypadku przekazujemy macierz cech (wektor różnych cech oraz wektory reprezentujące każdą z pojedynczych cech).

## Funkcja błędu

* + 1. Otrzymujemy wektor stanów po symulacji działania automatu
    2. Jeżeli symbol nie został rozpoznany, nie zwiększamy licznika rozpoznanych symboli, a w przypadku gdy został rozpoznany zwiększamy go o
    3. Obliczenia powtarzane są dla każdego symbolu
    4. Liczony jest procentowy błąd w klasyfikowaniu symboli

# Format uruchomieniowy programu