

UNIVERSITATEA POLITEHNICA DIN BUCUREȘTI

FACULTATEA DE ȘTIINȚE APLICATE

Matematică și Informatică Aplicată în Inginerie

PROIECT DE DIPLOMĂ

**CONDUCĂTOR ȘTIINȚIFIC,
Lect.univ.dr. Iuliana MUNTEANU**

**ABSOLVENT,
Robert Daniel SOARE**

București

2021



UNIVERSITATEA POLITEHNICA DIN BUCUREȘTI
FACULTATEA DE ȘTIINȚE APLICATE
Matematică și Informatică Aplicată în Inginerie



Aprobat Decan,
Prof.dr. Emil PETRESCU

PROIECT DE DIPLOMĂ

Învățarea automată pentru un agent într-un mediu 2D

CONDUCĂTOR ȘTIINȚIFIC,
Lect.univ.dr. Iuliana MUNTEANU

ABSOLVENT,
Robert Daniel SOARE

București

2021

Cuprins

Introducere	4
1 Învățare automată	7
1.1 Istoric	7
1.2 Clasificare	8
1.3 Industrie	9
1.4 Programe software pentru dezvoltare	10
1.5 Big Data	10
2 Rețele neuronale artificiale	13
2.1 Introducere	13
2.2 Structură	13
2.3 Funcții de activare și metode de optimizare	16
2.4 Tensorflow	16
3 Metode de învățare	19
3.1 Lanț și proces Markov	19
3.2 Ecuația Bellman	20
3.3 Q-Learning	20
4 Aplicație	21
4.1 Introducere	21
4.2 Structură	22
4.3 Simulator	23
4.4 Interfață	25
4.5 Agent	25
4.6 Model de învățare	26
Concluzii finale	26
Bibliografie	28
Index	29

Introducere

Învățarea automată a devenit un subiect de interes din ce în ce mai important, această fiind utilizată în vaste domenii, precum: industria auto, alimentară, agricolă, bancară, aerospațială și mai cu seamă în industria tehnologiei informației. Unul din rolurile ei cele mai importante constă în analiza și clasificarea datelor, predicția unor evenimente în baza unor fapte deja întâmplare, crearea unui profil virtual pentru un grup de utilizatori, etc.

Datorită marelui conectivități dintre oamenii din ziua de astăzi; sistemele politice, economice și relațiile interumane au devenit extrem de complexe. Totul a devenit interconectat. O idee a unui singur individ poate fi transmisă pe tot globul pământesc, aceasta idee putând afectând milioane de oameni în diverse locuri și a cărui impact politic și economic poate fi greu de estimat. De asemenea, un incident economic local, un dezastru natural, sau un conflict politic dintre două țări pot avea efecte devastatoare asupra economiei globale și a structurii geopolitice curente.

Fiecare eveniment din ziua de astăzi are o influență mai mică sau mai mare asupra acestei mari rețele de sisteme ale civilizației umane. Întrebarea naturală la această dilemă este: putem face o estimare asupra acestor evenimente și ale cazurilor lor speciale? Se poate, și asta datorită faptului că multe evenimente sunt monitorizate și înregistrate, precum: tranzacțiile bancare, documente legislative și juridice, vremea, traseele și destinațiile mașinăriilor de transport marfă (automobile, avioane, vapoare), discursuri și opinii în rețele sociale, date medicale din dispozitive inteligente (telefoane smart, ceasuri și brățari smart), date provenite din simulări virtuale sau experimente.

Tot acest mare volum de informații și metodele de manipulare, stocare întră în așa numita categorie *Big Data*. Analiză acestui volum imens de date devine o sarcină foarte dificilă și laborioasă în cazul metodelor convenționale de analiză a datelor folosind statistică clasică. În esență, învățarea automată se folosește atât de teoria clasică cât și de noile descoperiri în calculul numeric pentru a crea modele matematice dinamice care pot acumula cunoștințe și acționa în baza lor folosind toate datele pe care le primește ca set de învățare.

În această lucrare se va analiza cum algoritmi de învățare automată pot fi folosiți în crearea unui agent autonom care să îndeplinească sarcinile într-un spațiu 2-dimensional. Problema constă în rezolvarea unui traseu de tip matrice în care agentul trebuie să ajungă la destinație fără a produce un accident.

Analiza se va face cu ajutorul unei aplicații web interactive în care vom simula mediul

nostru 2-dimensional reprezentat de un labirint și care ne va permite analiza datelor furnizate de către agent în timpul sesiuni de antrenament pentru determinarea eficienței și fiabilității algoritmilor.

În primul capitol este descris termenul de învățare automată, care sunt subdomeniile sale, cum este folosit în industrie.

În al doilea capitol este o mică introducere pentru rețelele neuronale artificiale.

În al treilea capitol vor fi prezentați câțiva algoritmi de învățare, iar al patrulea descrierea aplicației.

În al patrulea capitol se va descrie structura aplicației web, a componentelor sale și modul cum acestea interacționează.

Capitolul 1

Învățare automată

1.1 Istoric

Învățarea automată este o ramură a inteligenței artificiale care se ocupă cu studiul tehnicilor și metodelor prin care se oferă unui calculator abilitatea de a învăța. Prin învățare ne referim la posibilitatea de a oferi o decizie în baza unor cunoștințe deduse din experiențe anterioare.

Multe tehnici din învățarea automată au la bază modelul de interacțiune al neuronilor, descris de către Donal Hebb în cartea sa *The Organization of Behavior* [1]. Termenul de învățare automată (în engleză *machine learning*) a aparut în anul 1953, dat de Arthur Samuel, creatorul unui program de jucat checker, capabil să ia decizii bazate pe experiențele anterioare [3]. În anul 1957, Frank Rosenblatt crează Perceptron-ul - utilizat în crearea unui calculator capabil să recunoască forme într-o imagine - folosindu-se de observațiile din lucrările lui Donald Hebb și Arthur Samuel. Perceptron-ul de unul singur are o putere destul de limitată, dar odată cu descoperirea utilizării sale în combinații de mai multe straturi a dat naștere la termenul de rețea neuronală.

De-a lungul timpului, acest domeniu a avut o evoluție înceată, un factor important fiind capacitățile limitate de procesarea ale calculatoarelor. Dar odată cu avansurile tehnologice, cercetarea în acest domeniu a început să fie din ce în ce mai activă, în ultimii ani culminând cu evenimente care au atras interesului publicului general, precum: IBM's Deep Blue, IBM's Watson, Google's Deepmind și Google's AlphaGo.

1.2 Clasificare

Fiind un domeniu foarte vast și cuprinzător, aceasta se împarte în 3 mari categorii:

- Învățare supervizată
- Învățare nesupervizată
- Învățare prin recompensă

În învățarea supervizată, procesul de antrenare se bazează pe analiza unor date formate din perechi de valori intrare-ieșire (set de date etichetat) pentru calibrarea funcțiilor de deducere. Este folosit pentru rezolvarea problemelor de clasificare.

Exemple de algoritmi:

- Support-vector machines
- Regresia liniară
- Regresia logistică
- Arbori de decizie
- Rețele neuronale
- Clasificator bayesian naiv

Pentru învățarea nesupervizată, procesul de antrenare constă în crearea unor modele interne de recunoaștere a unor tipare în urma analizei unui set de date neetichetat. Este deseori folosit în descoperirea similarităților și diferențelor într-un set de date.

Exemple de algoritmi:

- K-means clustering
- Autoencoders
- Analiza componentei principale
- Descompunerea valorilor singulare

În învățarea prin recompensă, procesul de antrenare constă în maximizarea unei funcții de recompensă, modelul calibrându-se astfel încât deciziile luate să ducă spre obținerea unei recompense cât mai mari.

Exemple de algoritmi:

- Monte Carlo
- Q-learning
- SARSA

- Deep Q Network
- Proximal Policy Optimization
- Deep Deterministic Policy Gradient
- Trust Region Policy Optimization

1.3 Industrie

Tot mai multe aplicații folosesc tehnici de învățare automată pentru optimizarea produselor, serviciilor și interacțiunilor cu utilizatorii. Cele mai notabile utilizări fiind:

- Algoritmi de căutare a știrilor în baza unor preferințe oferite explicit sau implicit de către utilizator.
- Reclame personalizate generate după profilele utilizatorilor.
- Sisteme de recomandări produse.
- Etichetarea obiectelor sau persoanelor în imagini, înregistrări audio sau video.
- Sisteme robotice autonome.
- Mașini autonome.
- Sisteme meteorologice
- Sisteme de detectare a fraudelor într-un sistem bancar.
- Clasificare și predicția evenimentelor.
- Optimizarea proceselor de producție a mărfurilor.
- Optimizarea procesului de antrenare pentru atleți.

Companiile sunt foarte interesate de modul cum interacționează și percep clienții produselor lor, ele încercând mereu să colecteze informații pentru despre modul cum sunt utilizate produsele în activitatea utilizatorului. Aceste campanii de colectare a datelor a devenit din ce în ce mai agresivă, marile companii software specializate în rețele sociale (Facebook, Twitter, Youtube, LinkedIn, Reddit) vând datele utilizatorilor în vederea oferirii unui profil al consumatorului pentru a stabili interesul pentru produs.

1.4 Programe software pentru dezvoltare

Interesul puternic pentru acest domeniu a venit în principal din partea marilor companii software și hardware, ele dezvoltând puternice biblioteci pentru procesarea datelor, crearea de rețele neuronale, algoritmi de învățare, etc. Pentru sprijinirea domeniului, aceste unelte sunt oferite după ca aplicații cu sursă deschisă (în engleză *open source*), având o licență deseori foarte permisibilă în vederea utilizării personale și comerciale.

Calitatea acestor unelte le-a făcut să devină un standard în industrie, atât comercială cât și academică.

Exemple de biblioteci sau aplicații software:

- Tensorflow - bibliotecă dezvoltată de către Google în vederea utilizării cu ușurință algoritmilor de învățare, cât și funcții utilitare pentru manipularea datelor.
- PyTorch - bibliotecă dezvoltată de către Facebook pentru protiparea aplicațiilor de viziune computerizate, procesarea limbajului natural, etc.
- ML.NET - bibliotecă dezvoltată de Microsoft pentru crearea rapidă a unor aplicații de procesare a datelor folosind algoritmi de învățare.
- scikit-Learn - bibliotecă care conține funcții statistice folosite pentru analiza datelor.
- Apache Spark - bibliotecă de aplicații destinate pentru procesarea unui volum foarte mare de date.
- Apache Kafka - aplicație care permite stocarea și distribuirea unui volum foarte mare de date în timp real către mai mulți consumatori.
- Caffe - bibliotecă pentru dezvoltare aplicațiilor pentru medii de lucru care nu dispun de o putere de procesare foarte mare, precum dispozitivele mobile.
- Keras - bibliotecă pentru dezvoltarea rețelelor neuronale
- H2O.ai - platformă de procesare și analiză a datelor pentru mediul comercial

1.5 Big Data

O componentă esențială pentru învățarea automată este gestionarea datelor care vor fi folosite și produse de către algoritmi algoritmi învățare. Această gestionare a informațiilor, de cele mai multe ori, va intra în cadrul domeniului de *Big Data*

Conform Uniunii Europene: „Big data se referă la volume de date colectate atât de mari și complexe încât este nevoie de noi tehnologii, cum ar fi inteligență artificială, pentru a le procesa. Datele provin din nenumărate surse diferite.” [5]

Volumul de date pe care omenirea îl produce crește de la an la an, ceea ce face analiza și înțelegerea datelor să fie o sarcină din ce în ce mai dificilă. Tot mai mulți oameni încep să

aibă acces la internet, iar numărul de dispozitive inteligente (smart phone, smart watch, smart TV) pe care un individ de dispune crește odată cu avansul tehnologic.

Principalele surse de proveniență ale acestor date sunt:

- Rețele sociale - mesaje, imagini create de utilizatori pentru ași exprima opinia la situația socială, economică și politică - datele pot fi utilizate pentru stabilirea unor tendințe sociale cu privire la activitatea și starea emoțională curentă și viitoare a oamenilor.
- Mediul și natura - date provenite de la sateliți și senzori pentru monitorizarea schimbărilor climatice - folosite pentru predicția posibilelor dezastre naturale cauzate de activitățile omului.
- Sector public - documente, certificate, atestate, adeverințe emise de către instituțiile publice - pot fi utilizate în eficientizarea serviciilor publice.
- Transport - date colectate prin GPS și de la diferiți operatori în domeniul transportului (transportul public, aeroporturi, gări) - pentru optimizarea rutelor și a curselor de transport.
- Sector Medical - fișe medicale ale pacienților - monitorizarea stării de sănătate a cetățenilor, utile pentru detectarea posibilelor amenințări de tip biologic.
- Internetul Lucrurilor (*Internet of Things*) - date provenite de la diverse aparate, precum: telefon, ceas, televizor, senzor de gaz, senzor de umiditate, camere video, etc. - utilizate la monitorizare activității individului cu scopul de a ușura anumite sarcini sau pentru a prevenii incidente.
- Sector industrial - rețele industriale de comunicații (senzori, magistrale de teren, rețele celulare), rapoarte economice - folosite pentru automatizare și îmbunătățirea produselor și a serviciilor.
- Sector bancar - tranzacții financiare, rapoarte - utilizate pentru detectarea fraudelor bancare, stabilirea ratelor la dobânzi, împrumuturi, schimb valutar, etc.

Toate aceste beneficii sunt importante pentru societatea din ziua de astăzi, companii mare concureaza pentru crearea de infrastructură și servicii pentru stocarea și examinarea datelor.

Exemple de servicii:

- Amazon Web Services - cel mai mare furnizor de servicii și infrastructură cloud din lume (având peste 200 de soluții software).
- Microsoft Azure
- Google Cloud Platform

- IBM Cloud
- Oracle Cloud
- Alibaba Cloud

Capitolul 2

Rețele neuronale artificiale

2.1 Introducere

O rețea neuronală artificială este un model computațional inspirat din structura și modul de funcționare al creierului biologic. Conexiunile dintre neuronii artificiali se asemănă sinapselor, fiecare neuron se conectează cu alt neuron prin intermediul unor muchii. Semnalul trimis prin aceste muchii este ponderat de niște parametri numiți ponderi sinaptice. Mai mulți neuroni grupați formează un strat, iar mai multe straturi formează o rețea.

Procesul de învățare presupune găsirea unor valori potrivite pentru ponderile sinaptice astfel încât procesarea semnalului de intrare să ofere rezultatul dorit.

2.2 Structură

Structura principală al unui neuron artificial este bazat pe modelul Perceptron-ului al lui Donald Hebb, modelul matematic fiind:

$$y = \varphi \left(\sum_{k=1}^n w_k * x_k + b \right)$$

,unde x este vectorul de intrare (*input vector*), y vectorul de ieșire (*output vector*), w ponderea sinaptică (*weight*), b deplasarea (*bias*) și φ este funcția de activare sau transfer (*activation function*).

Vectorul de intrare este format din numerele reale, aceste numere putând reprezenta: imagini, frecvențe, etichete codificate, valori provenite din senzori, etc. Ponderile sinaptice au rolul de a crește sau descrește puterea semnalului reprezentat de valorile vectorului de intrare. Funcția de activare preia semnalul ponderat și oferă o valoare specifică în baza acestuia. Deplasarea ajută la deplasarea semnalului ponderat pentru o mai bună aproximare necesară pentru îndeplinirea anumitor condiții ale funcției de activare.

Exemplul 2.2.1 *Un neuron artificial care acționează precum o poarta logică SAU (OR)*

pentru două numere binare are forma:

$$y = \varphi(x_1 + x_2 - 0.5)$$

,unde $x = \{x_i | x_i \in \{0, 1\}\}$, $y \in \{0, 1\}$, $w_1 = 1$, $w_2 = 1$, $b = -0.5$, iar funcția de activare este:

$$\varphi(u) = \begin{cases} 1 & u \geq 0 \\ 0 & u < 0 \end{cases}$$

Verificare. Pentru $x = [1, 0]$, avem $u = 1 + 0 - 0.5 = 0.5$ și $y = \varphi(u) = \varphi(0.5) = 1$ (același rezultat și pentru $x = [0, 1]$ - datorită proprietății de comutativitate a adunării). Pentru $x = [1, 1]$, avem $u = 1 + 1 - 0.5 = 1.5$ cu $\varphi(u) = \varphi(2) = 1$. Ultimul caz pentru $x = [0, 0]$, vom avea $u = 0 + 0 - 0.5 = -0.5$ cu $\varphi(-0.5) = 0$.

Observația 2.1 Fără funcția de activare, perceptronul acționează precum o funcție liniară. Prin utilizarea unei funcții de activare potrivite, puteam aborda mai ușor problemele neliniare, precum cele pentru clasificarea datelor în diverse categorii.

Un singur perceptron oferă doar o singură valoare de ieșire. Dacă dorim să avem mai multe valori de ieșiri trebuie să mai adăugăm perceptroni. Gruparea de neuroni artificiali se numește *strat*.

Structura unui strat format din perceptroni arată astfel în formă matriceală:

$$\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \varphi(u_1) \\ \varphi(u_2) \\ \vdots \\ \varphi(u_n) \end{bmatrix}$$

Rezultatele acestui strat pot fi transmise către un alt strat care poate avea o altă funcție de activare, astfel putem crea modele matematice mai complexe. Această înșiruire de straturi se numește *rețea*. Straturile intermediare sunt deseori referite ca *straturi ascunse*. Iar o rețea cu foarte multe straturi asunse poartă denumirea de *profundă (deep)*.

Rețele pot fi structurate și sub forma unui graf. Fiecare neuron fiind reprezentat de un nod, iar muchiile grafului sunt conexiunile dintre neuroni. Dacă graful suport nu conține cicluri, spunem că este uni-directional - o denumire uzuală peste acest tip de rețea este *feed-forward (FF)* (denumire pe care o vom folosi și în restul acestei lucrări). De asemenea, neuroni pot fi interconectați (graful suport conține cicluri), fapt care poate oferi rețelei mai multă putere de modelare. Acest tip de rețea este denumit în general *recurrent neural network (RNN)*

Rețelele neuronale artificiale pot fi considerate ca fiind "aproximatori universali" [6]:

Rețelele feed-forward multistrat sunt, în condiții generale ale funcției de activare ascunsă, aproximatori universali dacă dispun de un număr suficient de unități asunse.

De-a lungul anilor, au fost create foarte multe tipuri de rețele neuronale artificiale pentru a servi la rezolvarea de probleme din domenii dificile.

Exemple de tipuri de rețele:

- Feed Forward (FF)
- Deep Feed Forward (DFF)
- Radial Basis Network (RBF)
- Recurrent Neural Network (RNN)
- Long/Short Term Memory (LSTM)
- Markov Chain (MC)
- Deep Convolutional Network (DCN)
- Deconvolutional Network (DN)
- Support Vector Machine (SVM)
- Deep Belief Network (DBN)

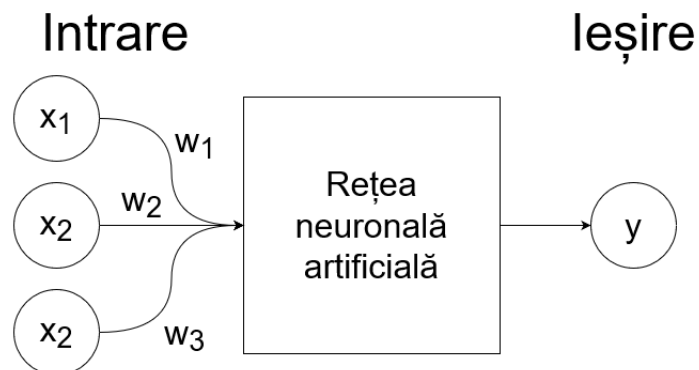


Figura 2.1: Structura unei rețele neuronale

2.3 Funcții de activare și metode de optimizare

Funcția de activare ajută rețeaua neuronală pentru învățarea de tipare complexe aflate în setul de date analizat. Alegerea unei funcții de activare este critică pentru performanța rețelei, în special cazul problemelor neliniare.

Unele din cele mai folosite funcții sunt:

$$\begin{aligned}
 \text{Identitate} \quad \varphi(x) &= x \\
 \text{Binary Step} \quad \varphi(x) &= \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \\
 \text{Logistic, Sigmoid} \quad \varphi(x) &= \frac{1}{1 + e^{-x}} \\
 \text{Rectified liniar unit (ReLU)} \quad \varphi(x) &= \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \\
 \text{Softplus} \quad \varphi(x) &= \ln(1 + e^x)
 \end{aligned}$$

Funcțiile de activare pot aproape orice funcție liniară sau neliniară, dar unele (precum cele enumerate anterior) oferă mai multe beneficii decât altele în contextul antrenării unei rețele.

Crearea unei rețele neuronale artificiale presupune alegerea tipurilor de straturi din care să fie compusă împreună cu funcțiile lor de activare. La început, ponderile sinaptice sunt de cele mai multe ori alese aleatoriu. Cea ce face ca aproximarea oferită de rețea să nu una foarte bună.

Acest lucru duce la problema de optimizare a rețelei (găsirea unor valori potrivite pentru ponderile sinaptice) astfel încât rezultatul aproximării să fie unul satisfăcător. Acest proces este referit uzual ca *antrenare* (*training*). Antrenarea este unul dintre cele mai dificile capitole al acestui domeniu, alegerea unui algoritm potrivit poate beneficia extraordinar în privința găsiilor unor valori optime pentru ponderile sinaptice.

Rețelele neuronale artificiale folosite în industrie (aplicații de recunoaștere a obiectelor în imagini; programe pentru traduceri, recunoașteri de voce) au o complexitate extraordinară atât din punct de vedere al arhitecturii, care poate constă dintr-un mix de diferite tipuri de rețele, cât și al nivelului imens de date (de ordinul milioanelor de terabiți). Antrenarea acestora poate dura câteva zile sau câteva luni, în cazuri rare fiind vorba de ani. Așadar, un algoritm de optimizare eficient are un rol crucial în acest proces.

2.4 Tensorflow

Tensorflow este o platformă dedicată dezvoltării modelelor de învățare automată. Această a fost creată inițial de către Google pentru a accelera dezvoltarea domeniului prin oferirea de programe ajutătoare pentru crearea rapidă a protipurilor. Datorită calității superioare a programelor de prototipare și a ușurinței de utilizare, acesta a devenit o platformă populară atât în mediul academic cât și industrial.

Datorită popularității, platforma a beneficiat de multe contribuții importante din partea marilor companii din domeniul IT și cel al semiconductoarelor, precum: PayPal, AMD, nVIDIA, Blomberg, Intel, IBM, Qualcomm, Uber, Arm, Twitter. De asemenea, platforma beneficiază de medii interactive de învățare, ideale pentru studenți sau profesioniști care doresc să dezvolte mici prototipuri de modele de învățare automată.

Exemple de programe care fac parte din platforma Tensorflow:

- Tensorflow Hub - biblioteca care găzduiește modele predefinite create de comunitate, precum: modele pentru clasificare imaginilor, analiza limbajului natural, generatoare de imagini
- Model Optimization - programe dedicate optimizării de modele
- TensorFlow Graphics - biblioteca care dispune de unelte pentru procesarea imaginilor
- TensorFlow Agents - biblioteca pentru dezvoltare agenților în cazul Învățării prin recompensă

În aceasta lucrare vom folosi aceste unelte pentru dezvoltarea unui model pentru agentul care va parcurge labirintul.

Capitolul 3

Metode de învățare

3.1 Lanț și proces Markov

La baza modelului de învățare pe care îl vom folosi în antrenarea agentului stau principiile fundamentale ale lanțului Markov și a procesului Markov. Proprietatea Markov afirmă că viitorul depinde numai de prezent și nu de trecut. Un lanț Markov este un model probabilistic care depinde numai de starea curentă pentru a prezice o stare viitoare. Așadar, un lanț Markov respectă proprietatea Markov. Trecerea de la o stare la alta se numește tranziție, iar probabilitatea ei poartă denumirea de probabilitate de tranziție.

De cele mai multe ori, lanțul Markov este reprezentat sub formă de graf orientat al cărui muchii reprezintă probabilitățile de tranziție dintre varfuri. Suma probabilităților de tranziție ale unui varf către celelalte varfuri este mereu 1.

În cazul agentului nostru, dacă ne-am imagina traseul ca fiind un lanț Markov, atunci pozițiile din trase ar fi varfurile grafului, iar mișcările agentului ar reprezenta muchiile. Putem asocia fiecărei mișcare o probabilitate, iar parcurgerea grafului reprezintă un posibil drum către obiectiv.

Așadar, dacă ne-am dori ca agentul să folosească această idee pentru stabilirea unui drum pentru rezolvarea obiectivului trebuie să stabilim o strategie de parcurgere a grafului. O strategie simplă ar fi una de tip *Greedy*, și anume agentul alege mereu acțiunea/muchia cu probabilitatea cea mai mare. Pentru ca această strategie să aibă succes trebuie ca valorile probabilităților acțiunilor să conducă către obiectiv în urma parcurgerii grafului.

Peste lanțul Markov putem contrui un model matematic pentru modelarea procesului de decizie al agentului. Acest model conține următoarele elemente:

- Mulțimea stărilor (S)
- Mulțimea acțiunilor (A)
- Probabilitatea de tranziție dintr-o stare în alta pentru o acțiune ($P_{ss'}^a$)
- Probabilitatea de a primi o recompensă în urma tranziției ($R_{ss'}^a$)

- Factor de atenuare, pondere care exprimă importanța recompenselor imediate și viitoare (γ)

Pentru a controla comportamentul agentului în mediul de lucru, vom folosi un sistem de recompense pentru fiecare decizie luată. Dacă dorim ca agentul să evite anumite situații precum luare de acțiuni care conduc la coliziuni cu anumite obiecte, vom asocia acestor decizii recompense negative. În contrast, dacă dorim ca agentul să facă anumite acțiuni care duc la îndeplinirea obiectivelor, acestor decizii li se vor asocia recompense pozitive.

Așadar, dorim ca pe parcursul simulării agentul să acumuleze cât mai multe recompense pozitive și să se evite pe cât mai mult posibil pe cele negative. Acestu lucru îl vom numi optimizare, iar procesul de antrenament implică modificarea rețelei neurale astfel încât acțiunile rezultate din datele de intrare să maximizeze recompensele acumulate.

3.2 Ecuția Bellman

3.3 Q-Learning

Capitolul 4

Aplicație

4.1 Introducere

Pentru aplicația în care vom implementa algoritmul de învățare ne dorim să avem disponibile următoarele funcționalități:

- Vizualizare grafică pentru mediul simulat - Afișarea labirintului sub forma unei poze
- Grafice pentru analizelor datelor opțiune din timpul sesiunilor de antrenament
- Elemente interactive care să ne permită modificare de parametri interni în algoritmi
- Tabele care să afișeze informații/date care urmează să fie procesate și rezultatele acestora

Având în vedere cerințele menționate anterior, platforma care ne poate permite în o dezvoltare rapidă asupra unei interfeței interactive foarte bogate este cea a aplicațiilor destinate web-ului.

Pentru contruirea aplicației vom folosi următoarele tehnologii:

- Svelte - program pentru construirea aplicațiilor web care va reprezenta interfața interactivă dedicată utilizatorului [7]
- Tensorflow.js - bibliotecă dedicată pentru contruirea și antrenarea rețelelor neuronale pentru aplicații web [9]
- Echarts - bibliotecă pentru contruirea de graficelor dedicate vizualizării de date [8]
- Konva - bibliotecă grafică pentru generarea imaginilor [10]

Elementele de bază vor fi construite folosind Svelte, acestea fiind: butoane, câmpuri de introducere a datelor, tabele pentru prezentarea informațiilor, containere pentru poziționare. Construirea labirintului se va face prin intermediul lui Konva, care ne permite și modificare rapidă a imaginii în funcție de starea internă a simulatorului. Datele provenite din

sesiunile de antrenament vor fi afișate imediat după fiecare sesiune cu ajutorul graficelor generate de Echarts.

Algoritmii de învățare vor fi construiți folosind funcțiile oferite de către Tensorflow, acestea oferind următoarele capabilități: crearea de straturi cu diverse funcționalități, funcții de activare, algoritmi de optimizare, metode utilitare de salvare și de creare a unor modele secvențiale formate din mai multe straturi de neuroni.

4.2 Structură

Labirintul este de forma unei matrici, fiecare celulă îndeplinește un anumit rol: drum, obstacol, ieșire, etc. Clasa principală dedicată definirii structurii de reprezentare a labirintului este denumită **Board**. Această clasă definește mediul simulat și interacțiunile disponibile pentru agent.

Pentru codificare vom avea următoarele reguli: spațiul liber va avea cifra 1, un obstacol cifra 2, iar pentru ieșire cifra 3. Pentru pozițiile agentului, la codificarea celulei de matrice se va adauga prefixul 1. Exemple de definiri ale mediului simulat prin codificare se pot observa în tabelul 4.1.

11	1	2	1	1	2	1	1	2	1	1	2	1	1	2
1	2	2	11	2	2	1	2	2	1	2	2	1	2	2
1	1	3	1	1	3	11	1	3	1	11	3	1	1	13

Tabelul 4.1: Exemple de codificări ale structurii labirintului

Sunt disponibile patru acțiuni pe care agentul poate să le ia: sus, jos, stânga și dreapta. Ca o acțiune să fie validă, aceasta trebuie să îndeplinească următoarea condiție: acțiunea nu trebuie să facă agentul să iasă din labirint atunci când se află la margine. Dacă se întâmplă acest lucru, clasa **Board** va pastra poziția agentului și va transmite faptul că mutare este invalidă celorlalte componente astfel încât acestea să poată stabili o recompensă astfel încât ca pe viitor agentul să evite astfel de situații.

Reprezentarea labirintului sub formă de imagine este dată de clasa **BoardUI**. Aceasta are rolul să creeze reprezentarea vizuală a labirintului folosind datele furnizate de către clasa **Board**. Această reprezentare poate fi observată în figura 4.1. Agentul este prezentat sub forma unui cerc albastru, spațiul liber sub forma unui pătrat gri, iar obiectivul cu un pătrat violet.

Clasă **BoardUI** este legată de clasa **Board** printr-un sistem reactiv de notificare, astfel încât orice schimbare care duce la modificarea stării labirintului (exemplu: mișcarea agentului) se propaga imediat către aceasta, astfel imaginea este modificată imediat cu noile date (figura 4.2).

Toate aceste imagini sunt generate folosind biblioteca Konva care ne permite atât generarea imaginii pentru afișare în pagina web cât și opțiune de interacțiune cu browserul

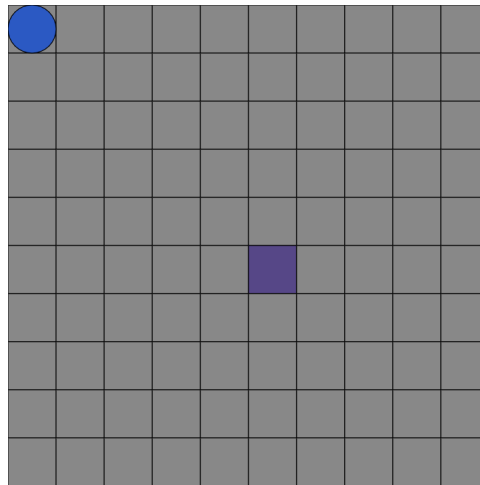


Figura 4.1: Reprezentarea vizuală a labirintului în starea inițială

care ne permit să transmitem evenimentele date de mouse (exemplu: click) către clasa **Board**.

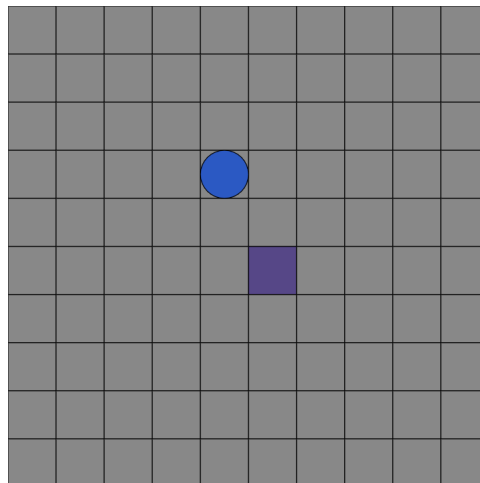


Figura 4.2: Reprezentarea vizuală a labirintului după o serie de acțiuni ale agentului

4.3 Simulator

Simulatorul acționează ca o interfață între agent și mediul reprezentat de labirint. Acesta are rolul să furnizeze informații către agent, precum: codificarea curentă a labirintului sau reprezentarea sa sub formă de imagine; dacă simularea este terminată; recompensa pentru fiecare acțiune luată.

Clasa **Simulator** este cea care definește structura modului de acționare a simulatorului. Acesta definește trei funcții principale, iar principiul lor de funcționare este inspirat

după standardul definit de către OpenAi Gym [2]. Așadar avem următoarele funcții:

- `reset` - această funcție aduce mediul simulat la starea inițială (figura 4.1) și returnează această stare sub forma dorită (codificare sau imagine).
- `step(acțiune)` - aceasta transmite acțiunea luată de agent către clasa **Board**, după acțiunea este procesată, clasa **Board** transmite înapoi valoare recompensei al acelei acțiuni dacă este validă, altfel doar anunța invaliditatea. După preluarea rezultatului, se decide care este recompensa și dacă simularea s-a încheiat în urma aceste acțiuni. Valorile noi stării, a recompensei și a semnalului de terminare sunt returnate la finalul evaluării.
- `actionSample` - această funcție returnează o acțiune aleatorie disponibilă în mediul simulat

```
1 class Env {
2     ACTIONS = ['UP', 'DOWN', 'RIGHT', 'LEFT']
3     invalidState = false
4     /**
5      *
6      * @param {Board} board
7      */
8     constructor(board) {
9         this.board = board
10    }
11
12    //
13    setAgentStartPosition(pos) {
14        this.board.playerDefaultPos = pos
15    }
16
17    //
18    step(action) {
19        this.invalidState = !this.board.move(this.ACTIONS[action])
20        return this.board.getBoardState(), this._getReward(), this._isDone
21    }
22
23    //
24    reset() {
25        this.board.playerReset()
26        return this.board.getBoardState()
27    }
28
29    //
30    actionSample() {
31        return Math.floor(Math.random() * this.ACTIONS.length)
32    }
33 }
```



```

34 //
35 _getReward() {
36     return (this.invalidState && !this.board.isOnExit() && -100) ||
    this.board.getPlayerCellValue()
37 }
38
39 //
40 _isDone() {
41     return this.board.isOnExit() || this.invalidState
42 }
43
44 //
45 clone() {
46     return new Env(this.board.clone())
47 }
48 }

```

Listing 4.1: Definirea clasei Simulator

4.4 Interfață

4.5 Agent

```

1 class Agent {
2     constructor(model) {
3         /**
4          * @type {tf.Sequential}
5          */
6         this.model = model || this.buildModel()
7     }
8
9     buildModel() {
10         const model = tf.sequential()
11
12         model.add(tf.layers.dense({ units: 10, inputShape: [10, 10],
    activation: 'relu' }))
13         model.add(tf.layers.flatten())
14         //model.add(tf.layers.dense({ units: 8, activation: 'relu' }))
15         //model.add(tf.layers.dense({ units: 16, activation: 'relu' }))
16         //model.add(tf.layers.dropout({ rate: 0.2 }))
17         // model.add(tf.layers.dense({ units: 32, activation: 'relu' }))
18         model.add(tf.layers.dropout({ rate: 0.2 }))
19         model.add(tf.layers.dense({ units: 4, activation: 'linear' }))
20         model.compile({ loss: 'meanSquaredError', optimizer: 'adam',
    metrics: ['accuracy'] })
21         model.summary()
22         return model
23     }
24 }

```

```
25   async fit(input, output) {  
26       await this.model.fit(tf.tensor3d([input]), output, { epochs: 5 })  
27   }  
28  
29   predict(input) {  
30       return this.model.predict(tf.tensor3d([input]))  
31   }  
32  
33   getAction(input) {  
34       const result = this.predict(input)  
35       return tf.argmax(result, 1).arraySync()[0]  
36   }  
37 }
```

4.6 Model de învățare

Concluzii finale

În această lucrare am analizat

Bibliografie

- [1] Hebb, D. O. The organization of behavior : a neuropsychological theory / D.O. Hebb, Wiley New York, 1949
- [2] Greg Brockman and Vicki Cheung and Ludwig Pettersson and Jonas Schneider and John Schulman and Jie Tang and Wojciech Zaremba, OpenAI Gym, CoRR, 2016
- [3] <http://infolab.stanford.edu/pub/voy/museum/samuel.html>
- [4] <https://www.ibm.com/cloud/learn/unsupervised-learning>
- [5] <https://www.europarl.europa.eu/news/ro/headlines/society/20210211ST097614/big-data-de>
- [6] Kurt Hornik, Approximation capabilities of multilayer feedforward networks, Neural Networks, 1991
- [7] <https://svelte.dev>
- [8] <http://echarts.apache.org/en/index.html>
- [9] <https://www.tensorflow.org/js>
- [10] <https://konvajs.org/docs/index.html>
- [11] Cormen, T.H., Leiserson, C.E., Rivest, R.L. *Introducere în algoritmi*, Cluj-Napoca, Editura Computer Libris Agora, 2000.
- [12] <https://www.mathworks.com/products/computer-vision.html>

Index

Învățare automată, 7, 13

capitol

 C2, 13

 C3, 19, 21

concluzii, 25

secțiune

 S1.1, 7