



Academia de Studii Economice din București
Școala Doctorală de Informatică Economică

Progrese în hibridizarea algoritmilor de inspirație biologică - Rezumatul tezei de abilitare -

Candidat

Conf.dr. Uscatu Cristian Răzvan

București, 2024

Teza de abilitare cu titlul *Progrese în hibridizarea algoritmilor de inspirație biologică* prezintă cele mai importante realizări ale autorului, conf. dr. Uscatu Cristian Răzvan, după obținerea titlului de doctor în anul 2007 și creionează principalele direcții de dezvoltare a carierei profesionale.

Teza este structurată în 3 secțiuni. Secțiunea I prezintă principalele rezultate ale cercetării publicate în articole reprezentative. Secțiunea II prezintă realizările profesionale, structurate pe direcțiile: activitate didactică, prestigiu profesional, publicații și proiecte de cercetare și indică creionează planurile de dezvoltare viitoare pe aceleași direcții. Ultima parte conține referințe bibliografice.

Secțiunea I prezintă propuneri de hibridizare a algoritmilor de inspirație biologică pentru rezolvarea problemelor complexe în domeniul procesării imaginilor și optimizării rețelelor neuronale. Problema de procesare a imaginilor optimizată prin utilizarea algoritmilor de inspirație biologică este alinierea imaginilor. Problema poate implica imagini captate de senzori diverși, în spectre diferite, la momente de timp diferite, din unghiuri diferite sau cu alte variații ale condițiilor de captare. Problema concretă abordată este alinierea imaginilor perturbate prin transformări geometrice. Optimizările propuse în cazul rețelelor neuronale se referă la îmbunătățirea parametrilor antrenabili și a hiper-parametrilor.

În capitolul 1 sînt prezentată o cercetare a literaturii în privința utilizării algoritmilor evolutivi în probleme de aliniere a imaginilor cu scopul de identificare a tendințelor în privința utilizării principalelor tipuri de elemente evolutive. Scopurile concrete urmărite sînt: identificarea celor mai recente lucrări în domeniu, publicate în ultimii 5 ani și indexate în principalele baze de date internaționale, disponibile liber pe internet sau prin intermediul rețelelor de cercetare la care este afiliată ASE; evidențierea cazurilor în care algoritmi evolutivi au fost folosiți cu succes pentru alinierea imaginilor; identificarea direcțiilor pe care se concentrează eforturile de cercetare; identificarea abordărilor utilizate în cercetare; identificare și analiza comparativă a utilizării indicatorilor de (dis)similaritate și a indicatorilor de performanță utilizați în analiza calității rezultatelor algoritmilor și pentru compararea algoritmilor; analiza comparativă a principalelor abordări evolutive: algoritmi de bază, clase de algoritmi și funcții calitate utilizate pentru evaluarea soluțiilor candidat. Concluziile cercetării arată că domeniul este unul foarte dinamic, cu noi algoritmi, operatori de variație și hibridizări propuse permanent, totul pe fondul dezvoltării rapide a tehnologiei care permite calcule rapide. Această evoluție tehnologică este fundalul pe care s-a extins utilizarea algoritmilor de căutare bazați pe populații, care necesită putere de calcul semnificativă.

În capitolul 2 este propus un nou algoritm hibrid pentru alinierea imaginilor. Calculele se fac pe variante binare ale imaginilor (care pot fi color sau monocrome). Modelul de perturbare luat în considerare este cel geometric, care implică rotație, translație și scalare (fie cu același factor pe ambele direcții, fie cu factor de scalare diferit pe fiecare direcție). Metodologia propusă implică hibridizarea algoritmului Firefly prin împărțirea populației în grupuri și aplicarea de îmbunătățiri locale asupra cîte unui membru din fiecare grup. Îmbunătățirile locale sînt realizate prin utilizarea strategiei evolutive cu doi membri. Numărul de grupuri variază la fiecare iterație, invers proporțional cu calitatea populației curente (măsurată ca fiind calitatea celui mai bun individ din populație), ca mecanism de prevenire a convergenței premature. Metodologia propusă a fost testată, iar experimentele au dovedit că produce rezultate mai bune decît metodele clasice, fiind în același timp mai eficientă (mai ales considerînd natura stocastică a algoritmului). Rata de succes (aliniere corectă) a fost de 100%, reușind să treacă peste problemele întîlnite de algoritmi clasici, inclusiv alinierea „cu capul în jos” care se poate produce în cazul utilizării metodei PAT.

În capitolul 3 metoda propusă anterior este dezvoltată și îmbunătățită prin adăugarea unui element nou: utilizarea de scale multiple ale imaginilor procesate. Pentru creșterea vitezei de procesare,

imaginile sînt reprezentate într-o scală redusă, care e utilizată pentru a calcula soluții candidat inițiale și a identifica direcții de căutare cu potențialul de a duce la optimul global. Acești candidați inițiali, care de fapt sînt aproximări grosiere ale posibilelor soluții, sînt amestecați cu candidați generați aleator pentru a forma populația inițială utilizată de algoritmul de optimizare bazat pe populații. Scala redusă este de asemenea utilizată pentru a înlocui parțial membrii populației inițiale în cadrul mecanismului de prevenire a convergenței premature. Rezultatele experimentale arată că metoda de aliniere propusă este precisă: 100% rată de succes, spre deosebire de algoritmi clasici, care pot avea probleme în cazul perturbărilor extreme. Metodologia propusă s-a dovedit de asemenea eficientă, accelerînd obținerea rezultatului de cel puțin două ori, comparativ cu varianta propusă anterior.

În capitolul 4, algoritmi de inspirație biologică sînt utilizați pentru a îmbunătăți predicțiile bazate pe serii de timp, prin utilizarea rețelelor neuronale. În urma cercetării exploratorii, rețelele neuronale de tip LSTM au fost alese ca fiind algoritmul cu cele mai bune perspective pentru acest tip de predicție, în combinație cu modelul NARX. Ținta optimizării e constituită de parametrii antrenabili ai rețelei neuronale. Rețeaua este antrenată folosind optimizatorul ADAM, producînd setul de valori pentru parametrii antrenabili. Acești parametri sînt optimizați prin intermediul unui algoritm din clasa strategiilor evolutive. Scopul principal este îmbunătățirea scorului F1 fără a afecta semnificativ metrica de eroare, ca urmare funcția de evaluare ca calității folosită de strategia evolutivă ia în considerare atît un indicator de calitate (scorul F1) cît și o metrică de eroare (MAPE). Strategia evolutivă aleasă este algoritmul ES2M, în care individul prelucrat este o concatenare a setului de parametri antrenabili ai rețelei neuronale. Testele au fost efectuate pe serii de timp notorii în privința dificultății predicției, și anume serii de date financiare. Au fost utilizate 3 seturi de date: ratele de conversie BTC-USD, ETH-USD și EUR-USD. Experimentele au arătat că metodologia propusă este superioară metodei de predicție bazate pe LSTM standard în ceea ce privește indicatorii POCID și scorul F1, în timp ce metrica de eroare MAPE nu e afectată semnificativ, ba chiar este îmbunătățită în unele cazuri.

În afară de îmbunătățirea predicțiilor pe serii de date de timp, un alt domeniu de utilizare a învățării automate o constituie clasificarea datelor, cu aplicații în diverse domenii. Un astfel de domeniu este diagnoza automată, bazată pe seturi de date medicale. Diagnoza automată poate fi folosită pentru a prelucra seturi mari de date și a detecta din timp posibile boli periculoase, precum afecțiunile cardiovasculare. În capitolul 5 astfel de metode de clasificare automată sînt explorate și analizate comparativ. Cele mai bune două metode rezultate în urma analizei au fost alese pentru hibridizare, urmărind îmbunătățirea performanțelor prin optimizarea hiper-parametrilor. Metodele de clasificare standard analizate sînt: regresia logistică (logistic regression), pădurile aleatoare (random forests), cei mai apropiați k vecini (k-nearest neighbors – KNN), mașinile suport vectoriale (support machine vectors – SVM, cu accent pe varianta soft margin SVM), și rețelele neuronale adînci (DNN) – cu accent pe rețelele tip LSTM. În urma testelor, metodele tip soft margin SVM și LSTM au obținut cele mai bune rezultate în privința preciziei și capacităților de generalizare. Aceste două metode au fost alese pentru optimizarea hiper-parametrilor. Pentru optimizarea metodei bazate pe SVM s-a ales folosirea unui algoritm din clasa strategiilor evolutive, iar pentru hibridizarea LSTM a fost ales estimatorul arbore structurat Parzen (structured-tree Parzen estimation). Metodele rezultate și propuse, numite MES-SVM și respectiv TPE-LSTM, au fost testate folosind baze de date publice cu date medicale. Ambele metode propuse au produs rezultate îmbunătățite față de metodele de bază.

Rezultatele promițătoare obținute de metodele propuse incită extinderea cercetării în acest domeniu.

Secțiunea II constă în două părți. Partea A sintetizează activitatea desfășurată în peste 25 ani de carieră universitară, structurată pe direcțiile activitate didactică, cercetare și publicații.

Pe direcția activitate didactică, retrospectiva sintetizează pașii succesivi ai carierei, cu toate cursurile și seminariile susținute în cadrul diverselor structuri ale Academiei de Studii Economice, atât în cadrul ciclului licență cât și în cadrul ciclului master: Facultatea de Cibernetică, Statistică și Informatică Economică (CSIE), Colegiul Universitar Economic București (CUEB), Colegiul Universitar Economic Giurgiu (CUEG), Programul Inter-universitar European Bulgaro-Român (BRIE). Sînt menționate și activitățile complementare: ghidarea și coordonarea studenților, participarea în diverse comisii și comitete legate de activitatea facultății și a universității, dezvoltarea de cursuri universitare etc.

Prestigiul profesional este evidențiat de diplomele primite de-a lungul carierei, participarea în asociații profesionale și comitete editoriale de revistă (indexate în baza de date internaționale) și invitații pentru recenzii de articole propuse pentru publicare în reviste științifice indexate ISI.

Direcția publicistică sintetizează publicațiile de pînă în prezent, care includ manuale universitare, cărți de specialitate, articole participante la conferințe internaționale și naționale, articole publicate în reviste indexate și diverse seminarii și citări ale lucrărilor publicate. Sînt subliniate articolele publicate în reviste indexate ISI și citările lucrărilor publicate de către articole publicate în reviste indexate ISI.

Partea B evidențiază planurile de dezvoltare a carierei, structurate de asemenea pe direcțiile activitate didactică, cercetare și publicistică. Principalele puncte urmărite sînt: continuarea și îmbunătățirea activităților curente, atât din punct de vedere curricular cât și al integrării tehnologiilor moderne; continuarea activității publicistice; deschiderea de noi oportunități prin crearea de cursuri noi sau extinse, explorarea de noi oportunități de cercetare și extinderea vizibilității prin noi canale de publicare.

Referințele bibliografice include articolele relevante și cărțile de referință utilizare în cercetare și citate în cele 5 articole care formează baza acestei teze. În total sînt citate 256 titluri.

Retrospectiva activităților didactice, de cercetare și publicare descrie un înalt nivel de experiență profesională. Împreună cu obiectivele formulate pentru dezvoltarea viitoare a carierei, dovedesc abilitatea de a organiza și conduce activități didactice, de ale sprijini prin realizarea de manuale și cărți de specialitate, de a aborda idei de cercetare interdisciplinare, de a coordona proiecte de cercetare și de a publica pe prin intermediul principalelor canale științifice.