

Algoritmi de învățare pentru rețele neuronale Clifford

Contribuția tezei

În această teză, am extins cei mai cunoscuți algoritmi de învățare din cadrul rețelelor neuronale de tip feedforward cu valori reale în domeniul **rețelelor neuronale de tip feedforward cu valori complexe**. Motivul pentru care am ales această temă este acela că am constatat, studiind literatura de specialitate, faptul că, cu excepția metodei gradient, a metodei gradient cu moment, a algoritmului resilient backpropagation și a algoritmului Levenberg-Marquardt, principalii algoritmi de antrenare a rețelelor cu valori reale nu au fost extinși în domeniul complex.

Ținând cont de această constatare, am prezentat deducerea completă a **metodelor gradient îmbunătățite** (respectiv algoritmi quickprop, resilient backpropagation, delta-bar-delta și SuperSAB), a **metodelor gradientilor conjugați** (cu actualizări Hestenes-Stiefel, Polak-Ribiere, Fletcher-Reeves și Dai-Yuan, și respectiv cu reporniri Powell-Beale, precum și două variațiuni ale actualizărilor Hestenes-Stiefel și Polak-Ribiere), a **metodei gradientilor conjugați scalați**, a **metodei Newton** (cu deducerea completă a algoritmului de calcul al **hessianei** funcției de eroare, care are și alte aplicații înafara acestui algoritm, și a **produsului hessianei cu un vector**), a **metodelor quasi-Newton** (respectiv metoda actualizării de rang unu și algoritmi Davidon-Fletcher-Powell, Broyden-Fletcher-Goldfarb-Shanno și one step secant) și a **metodei Levenberg-Marquardt**. Deducerea s-a făcut respectând toate particularitățile de calcul impuse de trecerea de la domeniul real la domeniul complex, avându-se în vedere obținerea unui grad maxim de generalitate, astfel încât algoritmi să poată fi utilizați într-un context cât mai larg. A fost folosit un tip de calcul special creat pentru derivarea în planul complex, și anume calculul Wirtinger, sau calculul $\mathbb{C}\mathbb{R}$.

Optsprezece algoritmi, dintre care **cincisprezece** apar aici pentru prima dată, au fost implementați în rețele neuronale cu valori complexe, care au fost testate pe mai multe benchmark-uri specifice domeniului și prezente extensiv în literatura de specialitate. Astfel, am prezentat rezultatele experimentale pentru aplicații sintetice, respectiv problema XOR, problema XOR extinsă, două probleme de aproximare a unor funcții complexe complet și trei probleme de aproximare a unor funcții complexe divizat, și pentru aplicații din lumea reală, respectiv egalizarea canalelor liniare și neliniare, predicția seriilor de timp liniare și neliniare și predicția direcției și vitezei vântului.

Rezultatele rulării de câte 50 de ori a fiecărui algoritm au fost prezentate în tabele și, pe o scară logaritmică, în figuri. Dinamica învățării a fost arătată în figuri. Se poate constata că din clasa metodelor gradient îmbunătățite, metoda resilient backpropagation și delta-bar-delta au avut, în general, cele mai bune performanțe. Din clasa metodelor gradientilor conjugați, metodele cu actualizări Hestenes-Stiefel, cu actua-

lizări Dai-Yuan, și respectiv cu reporniri Powell-Beale, au dat cele mai bune rezultate pe majoritatea problemelor. Dintre metodele quasi-Newton, metoda BFGS, a avut, în general, cea mai bună performanță.

Se poate constata de asemenea că toate clasele de algoritmi implementați au performanțe mai bune decât clasică metodă gradient. În general, metodele gradient îmbunătățite sunt cu un ordin de mărime mai bune în termeni de eroare medie pătratică decât metoda gradient, apoi urmează metodele gradientilor conjugați, iar în final, cu cel puțin un ordin de mărime mai bune decât acestea din urmă, sunt metodele quasi-Newton. Se mai poate constata însă în cazul anumitor probleme, tendința de overfitting în cazul metodelor quasi-Newton, ceea ce înseamnă că alegerea unuia sau altuia dintre algoritmi este foarte dependentă de problema de rezolvat, și este dificil de estimat apriori care dintre metode va da rezultatele cele mai bune.

Ca o concluzie generală a prezentului studiu, putem afirma că este de dorit și binevenită extinderea tuturor algoritmilor de optimizare a funcției de eroare pentru rețelele neuronale cu valori reale în domeniul complex, pentru că, în cazul anumitor probleme, ierarhiile dintre algoritmi în cazul real se inversează în cazul complex. Așa se întâmplă, de exemplu, cu metoda gradientilor conjugați cu actualizări Polak-Ribiere care în cazul real dă rezultate mai bune decât, spre exemplu, metoda gradientilor conjugați cu actualizări Fletcher-Reeves, iar în cazul complex, această ultimă metodă poate fi și cu un ordin de mărime în termeni de eroare medie pătratică mai bun decât algoritmul cu actualizări Polak-Ribiere.

Acest fapt ne încurajează ca pe viitor să considerăm extinderea și a altor algoritmi de învățare mai puțin cunoscuți din domeniul real în domeniul complex, deoarece, cu progresul aplicațiilor rețelelor neuronale cu valori complexe în lumea reală, este necesară o cât mai bună performanță a acestui tip de rețele. Datorită diferențelor fundamentale care există între domeniul real și complex, tot ceea ce se cunoaște în termeni de performanță în domeniul real, este posibil să nu se mai aplice în domeniul complex. În plus, acești algoritmi pot fi combinați cu tehnici specifice domeniului complex, pentru a le îmbunătăți și mai mult performanțele.

Ca o generalizare a rețelelor neuronale cu valori complexe, am dedus aceiași algoritmi pentru **rețelele neuronale Clifford**. Datorită interesului din ultimii ani pentru acest tip de rețele, am considerat oportun să prezentăm aceleași metode de învățare și pentru acest tip de rețele, ținând cont de faptul că doar metoda gradient a mai fost dedusă în altă parte în literatură, restul algoritmilor apărând aici pentru prima dată. Calculul diferențial folosit este cel cu derivate parțiale reale, care se poate particulariza și pentru rețelele neuronale cu valori complexe, dând formule diferite pentru algoritmi, care însă sunt echivalente cu cele deduse folosind calculul Wirtinger, sau calculul $\mathbb{C}\mathbb{R}$.

Astfel, am dedus formule pentru **metodele gradient îmbunătățite** (respectiv algoritmi quickprop, resilient backpropagation, delta-bar-delta și SuperSAB), **metodele gradientilor conjugați** (cu actualizări Hestenes-Stiefel, Polak-Ribiere, Fletcher-Reeves și Dai-Yuan, și respectiv cu reporniri Powell-Beale, și cele două variațiuni ale actualizărilor Hestenes-Stiefel și Polak-Ribiere), **metoda gradientilor conjugați scalați**, **metoda Newton** (cu deducerea completă a metodei de calcul al **hessianei** funcției de eroare și a **produsului hessianei cu un vector**), **metodele quasi-Newton** (respectiv metoda actualizării de rang unu și algoritmi Davidon-Fletcher-Powell, Broyden-Fletcher-Goldfarb-Shanno și one step secant) și pentru **algoritmul Levenberg-Marquardt**.

Motivul pentru care am decis să extindem aceiași algoritmi și pentru cazul Clifford este acela că, adesea, nu se cunoaște de la început care algebră Clifford este cea mai potrivită pentru a reprezenta datele dintr-o anumită aplicație. Astfel, am pus la dispoziția celor interesați o paletă largă de algoritmi, care pot fi aplicați pentru

rețele neuronale cu valori în diferite algebre Clifford, pentru a determina experimental care este cea mai potrivită configurație arhitectură-algoritm de învățare pentru fiecare problemă de interes. De exemplu, pentru cazul 2-dimensional există 2 algebre Clifford, pentru cazul 4-dimensional există 3 algebre Clifford, iar pentru cazul 8-dimensional avem 4 algebre Clifford cu care se poate experimenta.

O particularizare interesantă, cu multe aplicații în ultimul timp în literatura de specialitate, este aceea a **rețelelor neuronale cu valori cuaternionice**. Prin urmare, algoritmi mai sus propuși se pot particulariza ușor pentru algebra Clifford a cuaternionilor, de dimensiune 4. Pe viitor, implementări ale acestor algoritmi pentru rețelele neuronale cu valori cuaternionice promet să dea rezultate asemănătoare cu cele obținute de aceiași algoritmi în cazul complex. Literatura de specialitate din acest domeniu abia în ultimii trei ani a început să ia naștere, iar viitorul va putea da suficienți termeni de comparație pentru algoritmi propuși.

În această teză, am introdus de asemenea un algoritm ce permite calculul **mediei pe varietăți diferențiabile** fără a utiliza coordonate locale și derivate covariante pe varietatea respectivă. Acest algoritm, numit **algoritm de încorporare** (embedding algorithm) presupune liftarea funcției de cost inițiale la o varietate diferențiabilă care poate fi scufundată într-o varietate a riemanniană (spațiu euclidian) și construirea unui câmp de vectori definit pe spațiul ambient a cărui restricție la varietatea scufundată este câmpul de vectori gradient vector al funcției de cost liftate.

Astfel, pot fi calculate medii date de orice funcție de cost, indiferent de cât de complicată este expresia ei, pentru că algoritmul presupune doar calcule în coordonate carteziene. Au fost prezentate rezultate experimentale folosind funcții de tip L^p , ceea ce constituie o noutate în literatură, și în particular a fost făcută comparația între cazul L^2 și cazul L^4 , evidențind principalele deosebiri în ceea ce privește medierea folosind aceste două tipuri de funcții de cost.

O altă contribuție a prezentei teze este introducerea, în premieră după cunoștințele noastre, a unor generalizări ale rețelelor neuronale Clifford, și anume a **rețelelor neuronale cu valori matriciale**. În aceste rețele neuronale intrările, ieșirile, ponderile și bias-urile sunt matrici. Am considerat două cazuri, și anume **rețelele neuronale cu valori matrici pătratice**, care sunt definite pe algebra matricilor cu operațiile naturale de adunare și înmulțire, și respectiv **rețelele neuronale cu valori matrici antisimetrice**, adică rețele neuronale pe algebra asociată grupului de rotații.

În ambele cazuri, este introdusă metoda gradient de antrenare a rețelelor neuronale de tip feedforward. Ideea acestor rețele vine de la faptul că algoritmul propus pentru medierea matricilor ortogonale poate fi folosit pentru a genera seturi de date de antrenare pentru rețele neuronale, care pot astfel învăța medii de orice fel, permițându-ne să găsim media unor rotații fără a mai trece măcar prin calculele presupuse de algoritmul propus. Având valorile din grup, ele pot fi trecute în algebra asociată, date rețelei neuronale, iar rezultatele trecute înapoi în grup.

Rezultatele experimentale, pe trei probleme de aproximare a unor funcții sintetice în cazul rețelelor neuronale cu valori matrici pătratice, respectiv pe două probleme de aproximare a unor funcții sintetice și pe transformări geometrice (translația, scalarea și rotația) în cazul rețelelor neuronale cu valori matrici antisimetrice sunt promițătoare pentru viitorul acestor tipuri de rețele, care pot fi folosite pentru a rezolva probleme n -dimensionale pe care alți algoritmi cu valori reale și/sau Clifford nu au putut să le învețe, sau au avut performanțe slabe.

Structura tezei

Teza este structurată după cum urmează:

- Capitolul 2 face o prezentare sintetică a **rețelelor neuronale cu valori complexe**, a **rețelelor neuronale cu valori hiperbolice**, a **rețelelor neuronale cu valori cuaternionice** și a **rețelelor neuronale cu valori în algebre Clifford**, începând cu descrierea respectivelor sisteme de numere, și continuând cu principalele caracteristici ale rețelelor feedforward de tipurile respective.
- Capitolul 3 prezintă câte o secțiune pentru fiecare dintre metodele de învățare nou propuse pentru **rețelele neuronale cu valori complexe**, începând cu clasică *metodă gradient*, continuând cu *metodele gradient îmbunătățite*, *metodele gradientilor conjugați*, *metoda gradientilor conjugați scalați*, *metoda Newton cu calculul exact al hessianei* și *al produsului hessianei cu un vector*, *metodele quasi-Newton* și terminând cu *metoda Levenberg-Marquardt*. Pentru a deduce acești algoritmi, am folosit *calculul Wirtinger*, sau *calculul $\mathbb{C}\mathbb{R}$* , care este utilizat pe larg în literatura de specialitate pentru a extinde algoritmi de învățare din planul real în planul complex.
- Deoarece pentru **rețelele neuronale Clifford** nu există ceva asemănător calculului Wirtinger, am folosit calculul obișnuit cu derivate parțiale pentru a deduce *aceiași algoritmi* ca mai sus și pentru acest tip de rețele în Capitolul 4. Datorită generalității acestui tip de rețele, algoritmi deduși se pot particulariza ușor pentru *rețele neuronale cu valori hiperbolice*, pentru *rețele neuronale cu valori cuaternionice*, și chiar pentru rețele neuronale cu valori complexe, fiind o alternativă de exprimare a algoritmilor deduși în capitolul precedent.
- Capitolul 5 prezintă *metoda gradient* pentru antrenarea **rețelelor neuronale** de tip feedforward **cu valori vectori tridimensionali**, **cu valori vectori n -dimensionali**, și respectiv **cu valori în algebra matricilor pătratice** și **în algebra matricilor antisimetrice**. În același capitol se prezintă *algoritmul de încorporare* dezvoltat pentru a ușura calculul mediilor date de diferite funcții de cost pe varietăți riemanniene, și în particular **calculul mediilor pe grupul matricilor ortogonale**.
- Capitolul 6 prezintă **rezultatele experimentale** ale aplicării algoritmilor propuși pentru cele mai cunoscute benchmark-uri folosite în literatura de specialitate, atât pentru aplicații sintetice, cât și pentru aplicații din lumea reală, în cadrul *rețelelor neuronale cu valori complexe*. Aplicațiile sintetice includ problema XOR, problema XOR extinsă și aproximarea a două funcții complexe complet și a trei funcții complexe divizat. Aplicațiile din lumea reală sunt egalizarea cu canal liniar și respectiv neliniar, predicția seriilor de timp liniare și respectiv neliniare, precum și predicția direcției și vitezei vântului. Tot în acest capitol sunt prezentate aplicații sintetice pe probleme de aproximare de funcții ale metodei gradient pentru *rețelele neuronale cu valori matrici pătratice*, respectiv *cu valori matrici antisimetrice*, precum și rezultatele experimentale ale *medierii pe grupul matricilor ortogonale*.
- În final, Capitolul 7 prezintă **concluziile** prezentei teze.

Articole publicate sau comunicate

Rezultatele obținute în teză au fost valorificate în următoarele articole, dintre care 7 au fost publicate, iar unul este comunicat:

1. P. Birtea, D. Comănescu, and C.-A. Popa. Averaging on manifolds by embedding algorithm. *Journal of Mathematical Imaging and Vision*, 49(2):454 - 466, June 2014. (ISI Impact factor 2,33)
2. C.-A. Popa. Enhanced Gradient Descent Algorithms for Complex-Valued Neural Networks. *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. pages 272 - 279. IEEE, September 2014. (IEEE, ISI Proceedings)
3. C.-A. Popa. Exact Hessian Matrix Calculation for Complex-Valued Neural Networks. *International Workshop on Soft Computing Applications (SOFA)*. July 2014. (SpringerLink, ISI Proceedings)
4. C.-A. Popa. Scaled Conjugate Gradient Learning for Complex-Valued Neural Networks. *International Conference on Soft Computing (MENDEL)*. June 2015. (SpringerLink, ISI Proceedings)
5. C.-A. Popa. Matrix-Valued Neural Networks. *International Conference on Soft Computing (MENDEL)*. June 2015. (SpringerLink, ISI Proceedings)
6. C.-A. Popa. Quasi-Newton Learning Methods for Complex-Valued Neural Networks. *International Joint Conference on Neural Networks (IJCNN)*. July 2015 (IEEE, ISI Proceedings)
7. C.-A. Popa. Lie Algebra-Valued Neural Networks. *International Joint Conference on Neural Networks (IJCNN)*. July 2015 (IEEE, ISI Proceedings)
8. C.-A. Popa. Conjugate Gradient Algorithms for Complex-Valued Neural Networks. *International Conference on Neural Information Processing (ICONIP)*. November 2015 (SpringerLink, ISI Proceedings) – comunicat

Primul articol, publicat în jurnalul *Journal of Mathematical Imaging and Vision*, prezintă algoritmul de încorporare și este scris împreună cu doi matematicieni, P. Birtea și D. Comănescu, autorii apărând în ordine alfabetică, după cutuma existentă în cadrul comunității matematice. Având un factor mare de impact, jurnalul este unul dintre cele mai importante din domeniul inteligenței artificiale, și mai ales din domeniul imaging-ului, în care se aplică în particular algoritmul de încorporare pentru medierea matricilor ortogonale.

Pentru restul articolelor, autorul acestei teze este unic autor. Astfel, articolele 2 și 3 au apărut la conferințele *SYNASC 2014* și *SOFA 2014*, ambele ținute la Timișoara, și valorifică doi algoritmi pentru rețele neuronale cu valori complexe. Cele două conferințe sunt specifice domeniului inteligenței artificiale, iar a doua chiar soft computing-ului, din care face parte și domeniul rețelelor neuronale.

Articolele 4 și 5 au apărut la conferința *MENDEL 2015*, ținută la Brno, Cehia, de asemenea o conferință specifică domeniului soft computing, cu oarecare prestigiu internațional. De remarcat faptul că articolul 5 este cel care introduce, în premieră pentru literatura de specialitate, rețelele neuronale cu valori matrici pătratică, pe când articolul 4 prezintă un alt algoritm pentru rețelele neuronale cu valori complexe.

Însă poate cele mai importante articole, care dau cea mai mare legitimitate tezei, sunt articolele 6 și 7, prezentate la conferința *IJCNN 2015*, ținută la Killarney, Irlanda. Aceasta este cea mai importantă conferință în domeniul rețelelor neuronale, fiind sponsorizată de către International Neural Networks Society, și, în acest an, și de către European Neural Networks Society, pentru că ICANN, conferința acestei organizații, nu s-a ținut în 2015, tocmai pentru a face loc conferinței IJCNN, care are

loc în Europa. După cum am menționat în introducere, IJCNN a avut în ultimii ani o secțiune specială dedicată rețelelor neuronale cu valori complexe, și are în Program Committee unii dintre cei mai importanți cercetători în acest domeniu. Câtă vreme articolul 6 prezintă un algoritm pentru antrenarea rețelelor cu valori complexe, articolul 7 introduce, tot în premieră pentru literatura de specialitate, rețelele neuronale cu valori matrici antisimetrice.

Articolul 8 este dedicat unui alt algoritm pentru rețele neuronale cu valori complexe, fiind comunicat la *ICONIP 2015*.

În consecință, se poate spune că cercetarea făcută în această teză a fost validată de experți din domeniul imaging-ului, al inteligenței artificiale, al soft computing-ului, și mai ales din domeniul rețelelor neuronale cu valori complexe, fapt care îi oferă recunoaștere academică și vizibilitate internațională.