Universitatea din București
Facultatea de Matematică și Informatică

TEZĂ DE ABILITARE (REZUMAT)

Transferul cunoștințelor între vedere artificială, procesarea limbajului natural și biologie computațională

Radu Tudor Ionescu

București

2018

În cercetarea efectuată anterior, în cadrul doctoratului [1], am propus diverse metode și algoritmi care au avut la bază conceptul de a transfera cunoștințe între trei subdomenii diferite al informaticii, anume vederea artificială, procesarea limbajului natural și biologia computațională. Deși la prima vedere, aceste domenii de studiu sunt privite de majoritatea cercetătorilor ca domenii fără legătură între ele, o privire mai atentă poate dezvălui faptul că există multe concepte și principii comune. Mai mult, imaginile, documentele text sau secvențele ADN pot fi procesate cu tehnici similare.

După cum urmează a fi prezentat și în această teză de abilitare, idea generală de a trata imaginile și șirurile de caractere (documente text, secvențe ADN) într-o manieră similară se dovedește a fi foarte prolifică pentru o serie de probleme studiate în vederea artificială, procesarea limbajului natural și biologia computațională. Într-adevăr, una dintre metodele cele mai populare pentru recunoașterea obiectelor în imagini este inspirată de reprezentarea de tip *colecție-de-cuvinte (bag-of-words)* [2], care este utilizată pe scară largă în procesarea limbajului natural. În vederea artificială, reprezentarea analoagă se numește *colecție-de-cuvinte-vizuale (bag-of-visual-words)* [3, 4]. Această reprezentare constă în construirea unui vocabular de *cuvinte vizuale* prin gruparea descriptorilor de imagine folosind o tehnică de clustering, de exemplu algoritmul *k-mediilor (k-means)*. Reprezentarea de tip colecție-de-cuvinte-vizuale conduce la niveluri de performanță foarte ridicate pentru recunoașterea obiectelor [5], regăsirea imaginilor [6] sau alte probleme similare [7]. În general, prin adaptarea tehnicilor de procesare a șirurilor de caractere pentru analiza imaginilor sau invers, prin adaptarea tehnicilor de analiză a imaginilor pentru procesarea șirurilor de caractere, cunoștințele utile dintr-un domeniu pot fi transferate către celălalt domeniu. De fapt, multe descoperiri importante au avut la bază transferul cunoștințelor între domenii științifice diferte. Această teză de abilitare se încadrează în această direcție de cercetare, prezentând noi abordări precum și îmbunătățiri ale unor metode existente care se bazează pe transferul cunoștințelor.

În primul rând, prezentăm o *funcție nucleu (kernel function)* gândită să încorporeze informația spațială într-un mod eficient. Funcția kernel, introdusă în [8], este aplicată atât în problema recunoașterii claselor de obiecte din imagini (Capitolul 3) cât și în problema clasificării documentelor text după categorie (Capitolul 7), dovedind o îmbunătățire considerabilă a performanței în comparație cu reprezentarea de tip colecție-de-cuvinte standard și cu reprezentarea sub formă de piramidă spațială [9]. În al doilea rând, prezentăm o variantă diferită și mai performantă a modelului colecție-de-cuvinte (Capitolul 8), prin adaptarea modelului colecție-de-cuvinte-vizuale pentru lucrul cu documente text în loc de imagini, conform [10]. Adaptarea constă în înlocuirea descriptorilor de imagine necesari recunoașterii structurii obiectelor din imagini cu *scufundări vectoriale ale cuvintelor (word*

embeddings) [11] necesare recunoașterii structurii semantice a textelor. Modelul denumit *colecție-de-scufundări-vectoriale* este aplicat atât în clasificarea textelor după categorie sau după polaritatea opiniei cât și în notarea automată a eseurilor [12]. În al treilea rând, descriem o abordare pentru eliminarea valorilor aberante [13] prin ștergerea grupurilor cu mai puțini membri. Grupurile rezultă în urma aplicării algoritmului k-mediilor. Abordarea este aplicată atât în contextul detectării evenimentelor anormale din video (Capitolul 5), cât și în contextul dezambiguizării sensului cuvintelor (Capitolul 9), demonstrând performanțe de cel mai înalt nivel în ambele cazuri. În cel de-al patrulea rând, prezentăm o distanță pentru șiruri de caractere introdusă recent (Capitolul 6). Fiind gândită să se conformeze cu principii cât mai generale, dar fiind în același timp adaptată la secvențe ADN, noua distanță obține rezultate mai precise decât o serie de metode de ultimă oră pentru alinierea secvențelor ADN [14]. Totodată, prezentăm o adaptare a acestei distanțe, introdusă în [15], pentru problema recunoașterii gesturilor în secvențe video (Capitolul 4). Cele două distanțe pornesc de la aceeași sursă de inspirație, anume de la o măsură de disimilaritate pentru imagini [16] care la rândul său a fost inspirată de modul de calcul al distanței rang [17]. Nu în ultimul rând, atenția în teza de abilitare este îndreptată către abordările bazate pe învățarea nesupervizată. În acest sens, prezentăm o abordare pentru detectarea evenimentelor anormale din video (Capitolul 5) ce nu necesită date pentru antrenare. Abordarea constă în aplicarea tehnicii de *unmasking* [18], o metodă nesupervizată care a fost mai întâi propusă și utilizată pentru verificarea autorului unui text cu autor necunoscut. În același timp, prezentăm un algoritm nesupervizat, introdus în [19], pentru dezambiguizarea sensului cuvintelor (Capitolul 9) care este inspirat dintr-o metodă des utilizată pentru secvențializarea ADN-ului. În acest caz este vorba de transferul unei tehnici utilizate în biologia computațională către domeniul procesării limbajului natural. În concluzie, toate contribuțiile enumerate anterior vin să susțină ideea generală de a trata imaginile, documnentele text și secvențele ADN într-un mod similar, prin demonstrarea utilității practice a transferului de cunoștințe între domeniile care se ocupă cu analiza acestor tipuri de date.

Învățarea automată a devenit un subdomeniu larg al inteligenței artificiale prin faptul că are aplicabilitate în multe alte domenii precum vederea artificială, bioinformatica, regăsirea informației, procesarea limbajului natural, procesarea semnalelor, imagistică medicală și multe altele. În varietatea de abordări de ultimă oră din învățarea automată, se numără și metodele de *învățare profundă (deep learning)* și metodele de învățare pe bază de similaritate. Învățarea profundă se referă la antrenarea unor modele neuronale, care de obicei au mai multe straturi organizate secvențial, într-o manieră completă, de la stratul de intrare până la stratul de ieșire. Învățarea pe baza similarității se referă la procesul de învățare bazat pe similaritatea între toate perechile de exemple de antrenare. Procesul de învățare bazat pe similaritate poate fi atât supervizat cât și nesupervizat, iar relația între perechile de

exemple poate fi o funcție de similaritate, o funcție de disimilaritate sau o măsură de distanță. Cele două tipuri de metode de învățare sunt prezentate în Capitolul 2 al tezei, acestea fiind folosite în diverse abordări prezentate în Capitolele 3, 4, 5, 6, 7, 8, 9.

Pe de o parte, teza de față prezintă două metode, una nesupervizată [18] și alta supervizată [13], pentru detectarea evenimentelor anormale din video, în Capitolul 5. Ambele metode se bazează pe trăsături care modelează postura obiectelor. Aceste trăsături sunt extrase folosind rețele neuronale convoluționale antrenate folosind învățarea profundă. Pe de altă parte, teza de față studiază o serie de abordări ce folosesc învățarea bazată pe similaritate, cum ar fi metoda celor mai apropiați vecini, metode nucleu și algoritmi de clusterizare. Un model de tipul celor mai apropiați vecini bazat pe o nouă măsură de distanță [15] pentru secvențe temporale este prezentat în Capitolul 4. Metoda celor mai apropiați vecini este utilizată pentru recunoașterea limbajului semnelor din video, obținând rezultate foarte bune. Metodele nucleu sunt în utilizate în scopul soluționării mai multor probleme studiate în această teză. În Capitolul 3, este prezentată o funcție nucleu pentru histograme de cuvinte vizuale. Funcția kernel [8] îmbunătățește performanța pentru problema recunoașterii obiectelor din imagini prin utilizarea informației spațiale într-un mod eficient. Totodată, sunt prezentate mai multe funcții kernel bazate pe reprezentarea de tip piramidă spațială. Toate aceste funcții kernel, sunt utilizate pe de o parte pentru recunoașterea obiectelor în imagini (Capitolul 3) și pe de altă parte pentru clasificarea textelor după categorie și după polaritatea opiniei (Capitolul 7). Din varietatea de algoritmi de clusterizare, algoritmul k-mediilor este cel mai des întâlnit în teza de abilitare. Algoritmul este utilizat atât pentru obținerea vocabularelor de cuvinte vizuale în problema recunoașterii obiectelor (Capitolul 3), cât și pentru obținera vocabularelor de scufundări vectoriale ale cuvintelor în problema clasificării textelor (Capitolul 8). De asemenea, algoritmul k-mediilor este utilizat pentru detectarea valorilor aberante atât în contextul detectării evenimentelor anormale din video (Capitolul 5) cât și în contextul dezambiguizării sensului cuvintelor la nivel de document (Capitolul 9).

Se poate observa cu ușurință că problemele abordate în această teză se împart în două domenii separate, anume vederea artificială și procesarea șirurilor de caractere. Chiar dacă teza de abilitare explorează ambele domenii, este de remarcat faptul că metodele studiate în teză obțin rezultate de cel mai înalt nivel, o parte dintre aceste metode fiind prezentate în conferințe de top din cele două domenii, anume ICCV [17], ACL [12] și EACL [18]. Capitolul 10 încheie teza de față cu o serie de concluzii generale și cu câteva direcții ce vor fi explorate în viitor.

Referințe bibliografice:

[1] Ionescu, Radu Tudor. Machine Learning in Computer Vision and String Processing. PhD Thesis, University of Bucharest, 2013.

[2] Manning, Christopher D. and Schutze, Hinrich. Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, MA, USA, 1999.

[3] Csurka, Gabriella, Dance, Christopher R., Fan, Lixin, Willamowski, Jutta, and Bray, Cdric. Visual categorization with bags of keypoints. In Proceedings of Workshop on Statistical Learning in Computer Vision at ECCV, pp. 1–22, 2004.

[4] Leung, Thomas and Malik, Jitendra. Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. International Journal of Computer Vision, 43(1):29–44, 2001.

[5] Zhang, Jian, Marszalek, Marcin, Lazebnik, Svetlana, and Schmid, Cordelia. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. International Journal of Computer Vision, 73(2):213– 238, 2007.

[6] Philbin, James, Chum, Ondrej, Isard, Michael, Sivic, Josef, and Zisserman, An- drew. Object retrieval with large vocabularies and fast spatial matching. In Proceedings of CVPR, pp. 1–8, 2007.

[7] Ionescu, Radu Tudor, Popescu, Marius, and Grozea, Cristian. Local Learning to Improve Bag of Visual Words Model for Facial Expression Recognition. In Proceedings of ICML Workshop on Challenges in Representation Learning, 2013.

[8] Ionescu, Radu Tudor and Popescu, Marius. Have a SNAK. Encoding Spatial Information with the Spatial Non-alignment Kernel. In Proceedings of ICIAP, volume 9279, pp. 97–108. Springer LNCS, 2015.

[9] Lazebnik, Svetlana, Schmid, Cordelia, and Ponce, Jean. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In Proceedings of CVPR, volume 2, pp. 2169–2178, Washington, DC, USA, 2006.

[10] Butnaru, Andrei and Ionescu, Radu Tudor. From Image to Text Classification: A Novel Approach based on Clustering Word Embeddings. In Proceedings of KES, pp. 1784–1793, 2017.

[11] Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Gregory S., and Dean, Jeffrey. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS, pp. 3111–3119, 2013.

[12] Cozma, Mădălina, Butnaru, Andrei and Ionescu, Radu Tudor. Automated essay scoring with string kernels and word embeddings. In Proceedings of ACL, 2018.

[13] Ionescu, Radu Tudor, Smeureanu, Sorina, Popescu, Marius, and Alexe, Bogdan. Detecting abnormal events in video using Narrowed Motion Clusters. CoRR, abs/1801.05030, 2018.

[14] Dinu, Liviu P., Ionescu, Radu Tudor, and Tomescu, Alexandru I. A rank-based sequence aligner with applications in phylogenetic analysis. PLoS ONE, 9(8): e104006, 08 2014.

[15] Ionescu, Radu Tudor, Popescu, Marius, Conly, Christopher, and Athitsos, Vassilis. Local Frame Match Distance: A novel approach for exemplar gesture recognition. In Proceedings of EUSIPCO, pp. 788–792, 2017.

[16] Dinu, Liviu P., Ionescu, Radu Tudor, and Popescu, Marius. Local Patch Dissimilarity for Images. In Proceedings of ICONIP, volume 7663, pp. 117–126, 2012.

[17] Dinu, Liviu P. and Manea, Florin. An efficient approach for the rank aggregation problem. Theoretical Computer Science, 359(1–3):455–461, 2006.

[18] Ionescu, Radu Tudor, Smeureanu, Sorina, Alexe, Bogdan, and Popescu, Marius. Unmasking the abnormal events in video. In Proceedings of ICCV, pp. 2895– 2903, 2017.

[19] Butnaru, Andrei, Ionescu, Radu Tudor, and Hristea, Florentina. ShotgunWSD: An unsupervised algorithm for global word sense disambiguation inspired by DNA sequencing. In Proceedings of EACL, pp. 916–926, 2017.