

Vedecká rada Fakulty matematiky, fyziky a informatiky
Univerzity Komenského v Bratislave

RNDr. Pavol Vančo

Autoreferát dizertačnej práce

**Processing of Tree-Structured Data with Recursive
Self-Organizing Maps**

(Spracovanie stromových štruktúr pomocou rekurzívnych
samoorganizujúcich sa máp)

na získanie vedecko-akademickej hodnosti philosophiæ doctor

v odbore doktorandského štúdia:

9.2.1. Informatika

Bratislava 2010

Dizertačná práca bola vypracovaná v internej forme doktorandského štúdia na Katedre aplikovanej informatiky Fakulty matematiky, fyziky a informatiky Univerzity Komenského v Bratislave.

Predkladateľ: RNDr. Pavol Vančo
Katedra aplikovanej informatiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
Mlynská dolina
842 48 Bratislava

Školiteľ: doc. Ing. Igor Farkaš, PhD.
Katedra aplikovanej informatiky FMFI UK
Bratislava

Oponenti: doc. RNDr. Gabriela Andrejková, CSc.
Ústav informatiky, Prírodovedecká fakulta UPJŠ, Košice

prof. Ing. Vladimír Kvasnička, DrSc.
Fakulta informatiky a informačných technológií STU, Bratislava

doc. RNDr. Jiří Šíma, DrSc.
Ústav informatiky, Akademie věd ČR, Praha

Obhajoba dizertačnej práce sa koná o h
pred komisiou pre obhajobu dizertačnej práce v odbore doktorandského štúdia vymenovanou predsedom odborovej komisie

9.2.1. Informatika

na

Predseda odborovej komisie:
prof. RNDr. Branislav Rován, PhD.
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
Mlynská dolina
842 48 Bratislava

1 Úvod

V praxi prirodzene pracujeme z mnohými štruktúrami, ako sú napríklad vývoj finančného trhu, sekvencie DNA, obrázky alebo web stránky. Na spracovanie takýchto objektov je možné použiť neurónové siete, ktoré majú veľa pozitívnych vlastností, ako napríklad redundancia, robustnosť alebo možnosť spracovania zašumených dát.

Zaužívané spracovanie štruktúrovaných dát pomocou symbolických modelov je založené na spracovaní vrcholov. Tieto dáta sú uložené v báze znalostí a spracovávané pravidlami. Pravidlá sú vytvorené explicitne. Takéto spracovanie je síce efektívne, ale neumožňuje spracovanie zašumených dát a redundanciu tak, ako neurónové siete.

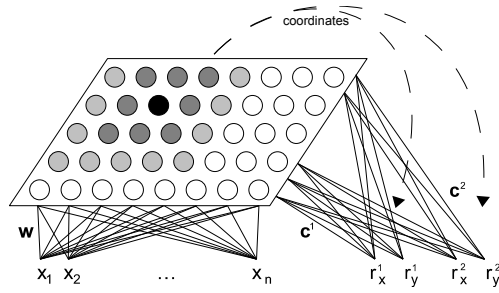
Predkladaná práca prezentuje možnosti spracovania sekvencií a stromových štruktúr neurónovými sieťami. Konkrétne sa zameriava na oblasť modelov, s učením bez učiteľa, nazývané samoorganizujúce sa mapy. Tieto boli nedávno rozšírené zo sekvencií na stromové štruktúry, ale ich komplexné porovnanie neexistovalo. Práca predkladá nové experimenty, ktoré slúžia na porovnanie modelov a rozoberá ich klady a zápory. Pokračuje použitím týchto modelov na XML dátach. Prezentuje výhody a nevýhody dávkového učenia a rozširuje existujúce dávkové učenie na použitie pre nové modely. Takisto poukazuje na možnosť využitia máp ako modelov pamäte a ukazuje možnosti extrakcie a rekonštrukcie uložených informácií.

2 Základné pojmy a označenia

V tejto časti prezentujeme tri modely rekurzívnych samoorganizujúcich sa máp (SOM). Pod modelmi rekurzívnych SOM rozumieme modely spracujúce stromové štruktúry. Viac o samoorganizujúcej sa mape je možné nájsť v [29, 61, 46, 62, 63]

2.1 SOMSD

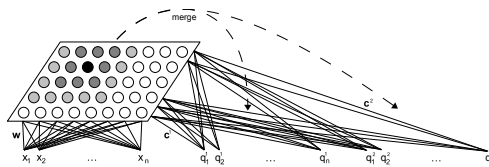
SOM pre štruktúrované dáta (SOMSD) [20] je model priamo vytvorený na spracovanie stromov. Tento model používa pozíciu (koordináty) víťazného neurónu ako kontext pre ďalší krok. Ako všetky modely rekurzívnych SOM, pre stromy s maximálnym vetvením k , je potrebných k kontextov. Obr. 1 zobrazuje model SOMSD pre binárne stromy.



Obr. 1: Model SOMSD – kontext pozostáva z koordinátov predchádzajúceho víťaza (\mathbf{r}).

2.2 MSOM

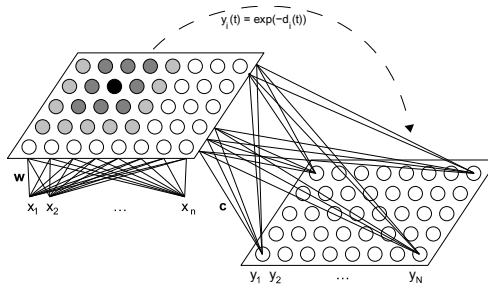
Model MSOM (Merge SOM) na rozdiel od modelu SOMSD používa ako kontext viac informácie o víťazovi, konkrétne hodnotu, ktorá vznikne spojením jeho váhových vektorov (obr. 2). Tento model bol pôvodne vytvorený na spracovanie sekvencií [67] a v tejto práci rozšírený na spracovanie stromov.



Obr. 2: Model MSOM pre stromové štruktúry.

2.3 RecSOM

Najzložitejší model, rekurzívna SOM (RecSOM) [77], bol takisto pôvodne vytvorený na spracovanie sekvencií (obr. 3). Tento model používa aktiváciu celej mapy ako kontext, a preto ma vysokú pamäťovú aj časovú zložitosť.



Obr. 3: Model RecSOM pre sekvencie.

3 Ciele dizertačnej práce

Základné ciele našej dizertačnej práce môžeme zhrnúť do štyroch bodov:

1. Navrhnuť a implementovať rekurzívnu verziu modelu MSOM.
2. Experimentálne porovnať existujúce rekurzívne samoorganizujúce sa mapy.
3. Preskúmať možnosti praktického využitia rekurzívnych samoorganizujúcich sa máp.
4. Navrhnuť dávkové učenie pre modely rekurzívnych SOM.

4 Výsledky práce a ich význam

V tejto časti predkladáme výsledky našej práce. V 4.1. časti sa zaoberáme výsledkami zodpovedajúcimi 1. cieľu. V časti 4.2 prechádzame výsledkami súvisiacimi s cieľom 2. Cieľ 3 je popísaný v časti 4.3 a cieľu 4 je venovaná časť 4.4.

V poslednej časti, 4.5, uvádzame stručné zhrnutie prezentovaných výsledkov v práci a ich význam.

4.1 Rozšírenie modelu MSOM

Model MSOM bol vytvorený na spracovanie sekvencií. V Hammer (2004) [28] bolo prezentované, že z dôvodu komutatívnosti pri výpočte kontextu tento model nemôže rozoznať synov v strome. Teda predpokladalo sa, že tento model bude schopný jednoznačne spracovávať len niektoré dátové množiny. Boli navrhnuté výpočtové postupy, ktoré upravujú správanie sa modelu.

Pri práci s týmto modelom sme zistili, že nie je dôvod k tomuto predpokladu. Matematickou úvahou sme dospeli k dôkazu, že tento model vie priamo rozoznávať synov v strome. Teoretický výsledok sme podporili aj jednoduchým ale jasným experimentom, ktorý naše závery potvrdil.

4.2 Porovnanie existujúcich modelov rekurzívnych SOM

Modely rekurzívnych SOM prezentované v tejto práci porovnávame na troch dátových množinách rôznej zložitosti. Za účelom porovnania definujeme a aplikujeme šesť kvantitatívnych mier, ktoré sa zameriavajú na rôzne vlastnosti natrénovaných máp.

Porovnávané modely sa líšia v spôsobe rozlišovania stromov a ich klasterizácie. SOMSD zdôrazňuje štruktúru pred obsahom, ale obsah sa stáva dôležitým v prípade, keď je štruktúra rovnaká. MSOM klasterizuje stromy podľa obsahu viac ako SOMSD, ale štruktúra je stále veľmi dôležitá. RecSOM vstupné dáta organizuje komplexne. Výsledná organizácia je založená na štruktúre aj obsahu. Čo sa týka jednoznačnosti výstupných reprezentácií, index víťaza stačí na rozlíšenie vstupov iba v prípade najjednoduchšej vstupnej množiny, keď je počet neurónov vyšší ako počet všetkých vstupných štruktúr.

4.3 Praktické využitie modelov rekurzívnych SOM

Na praktické využitie prezentovaných modelov sa sústreďujú dve kapitoly v práci – spracovanie štruktúrovaných dát z XML a extrakcia a rekonštrukcia dát.

V prvej sa sústreďujeme na použitie existujúceho formátu štruktúrovaných dát, XML, pri spracovaní modelmi rekurzívnych SOM. Dve dátové množiny sú vybrané ako príklad možností spracovania a použitia v praxi. V prípade prvej, jednoduchšej množiny, ukazujeme možnosti vizualizácie a v prípade druhej, komplexnej množiny, ukazujeme možnosti klasterizácie vstupov a objavovania znalostí.

V druhej z týchto kapitol prezentujeme možnosť extrakcie dát z natrénovaných modelov, a to buď pomocou tabuľky alebo pomocou doprednej neurónovej siete. Ak nie je možné získať úplné dáta, uvádzame niektoré metódy rekonštrukcie dát.

4.4 Dávkové učenie pre modely rekurzívnych SOM

Pre klasický, nerekurzívny, model ukazujeme výhody a nevýhody dávkového učenia. Pri nevýhodám prezentujeme konkrétne simulácie, keď je použitie dávkového učenia kontraproduktívne. V ostatných prípadoch je dávkové učenie vhodnejší a rýchlejší spôsob učenia.

Využitím existujúcich algoritmov dávkového učenia navrhujeme a implementujeme dva typy dávkového učenia pre rekurzívne samoorganizujúce sa mapy. Prvý typ umožňuje spracovať štruktúru ako celok. Druhý typ využíva spracovanie štruktúry ako celku a rozširuje ho na spracovanie celej dátovej množiny. Učenie teda nastáva až po prezentovaní celej dátovej množiny, pričom nezáleží na poradí prezentovaných vstupov.

Pomocou navrhnutého dávkového učenia je možné trénovať veľké mapy distribuovane, teda na viacerých počítačoch naraz.

4.5 Význam výsledkov práce

Táto dizertačná práca systematicky porovnáva modely rekurzívnych samoorganizujúcich sa máp. Jej hlavným prínosom je analýza modelov, ich správania, ich využitia pre rôzne typy dátových množín.

Medzi najdôležitejšie výsledky patrí dôkaz funkčnosti rekurzívneho modelu MSOM, bez nutnosti predspracovania alebo iných výpočtových metód. Týmto sa model MSOM zaradil medzi dôležité rekurzívne modely.

Takisto návrh dávkového učenia umožňuje distribuovaný výpočet počas učenia, a teda využitie počítačových klastrov na výpočty.

5 Zoznam použitej literatúry

- [1] Z. Bar-Joseph, E. Demaine, D. Gifford, and T. Jaakkola. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17:22–29, 2001.
- [2] G. Barreto, A. F. R. Araújo, and S. C. Kremer. A taxonomy of spatiotemporal connectionist networks revisited: The unsupervised case. *Neural Computation*, 15(6):1255–1320, 2003.
- [3] Ľ. Beňušková. Neurón a mozog. Prednáška Neurovedy I, 26.10.2000 v rámci celo-UK predmetu Kognitívne vedy, 2000.
- [4] M. Čerňanský and P. Tiňo. Comparison of echo state networks with simple recurrent networks and variable-length Markov models on symbolic sequences. In *17th International Conference on Artificial Neural Networks*, pages 618–627, 2007.
- [5] G. J. Chappell and J. G. Taylor. The temporal Kohonen map. *Neural Networks*, 6:441–445, 1993.
- [6] Y. Cheng. Convergence and ordering of Kohonen’s Batch Map. *Neural Computation*, 9(8):1667–1676, 1997.
- [7] R. Čihák. *Anatomie 3*. Grada Publishing, 2004.
- [8] F. I. Craik and R. S. Lockhart. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11:671–684, 1972.
- [9] F. I. Craik and E. Tulving. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3):268–294, 1975.
- [10] R. Diestel. *Graph Theory (Graduate Texts in Mathematics)*. Springer, Heidelberg, August 2005.
- [11] J. L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.
- [12] I. Farkaš and M. Pokorný. Processing tree-structured data with the linear RAAM neural network. Technical Report TR-2007-011, Comenius University in Bratislava, 2007.

- [13] I. Farkaš and P. Vančo. Spracovanie postupností symbolov pomocou rekurzívnych neurónových máp. *Kognice a umělý život*, 7:99–105, 2007.
- [14] J. A. Fodor and Z. W. Pylyshyn. Connectionism and cognitive architecture: a critical analysis. In S. Pinker and J. Mehler, editors, *Connections and Symbols*. MIT Press, Cambridge, Mass., 1988.
- [15] J. C. Fort, M. Cottrell, and P. Letremy. Stochastic on-line algorithm versus batch algorithm for quantization and self organizing maps. *Neural Networks for Signal Processing*, 9:43–52, 2001.
- [16] J. C. Fort, M. Cottrell, and P. Letremy. Advantages and drawbacks of the Batch Kohonen algorithm. *ESANN Proceedings*, 2002:223–230, 2002.
- [17] P. Frasconi, M. Gori, A. Kuechler, and A. Sperduti. *A Field Guide to Dynamic Recurrent Networks*, chapter From Sequences to Data Structures: Theory and Applications, pages 351–374. IEEE Press, Cambridge, CA, 2001.
- [18] P. Frasconi, M. Gori, and A. Sperduti. A general framework for processing of data structures. *IEEE Transactions on Neural Networks*, 9(5):768–786, 1998.
- [19] M. Gori, M. Mozer, A. C. Tsoi, and R. Watrous. Special issue on recurrent neural networks for sequence processing. *Neurocomputing*, 15(3–4):181–182, 1997.
- [20] M. Hagenbuchner, A. Sperduti, and A. C. Tsoi. A self-organizing map for adaptive processing of structured data. *IEEE Transactions on Neural Networks*, 14(3):491–505, 2003.
- [21] M. Hagenbuchner, A. Sperduti, and A. C. Tsoi. Contextual processing of graphs using self-organizing maps. In *13th European symposium on Artificial Neural Networks*, pages 399–404, 2005.
- [22] M. Hagenbuchner, A. Sperduti, and A. C. Tsoi. Contextual self-organizing maps for structured domains. In *ECML Workshop on Relational Machine Learning*, pages 46–55, 2005.
- [23] M. Hagenbuchner, A. Sperduti, A. C. Tsoi, F. Trentini, F. Scarselli, and M. Gori. *Lecture Notes in Computer Science 3977*, chapter Clustering XML documents using self-organizing maps for structures, pages 481–496. Springer-Verlag, 2005.

- [24] B. Hammer. *Learning with Recurrent Neural Networks*. Springer Lecture Notes in Control and Information Sciences 254. Springer, 2000.
- [25] B. Hammer. *Perspectives on learning symbolic data with connectionistic systems*, chapter Adaptivity and Learning, pages 141–160. Springer, 2003.
- [26] B. Hammer and B. J. Jain. Neural methods for non-standard data. In M. Verleysen, editor, *European Symposium on Artificial Neural Networks*, pages 281–292. D-side Publications, 2004.
- [27] B. Hammer, A. Micheli, A. Sperduti, and M. Strickert. A general framework for unsupervised processing of structured data. *Neurocomputing*, 57:3–35, 2004.
- [28] B. Hammer, A. Micheli, A. Sperduti, and M. Strickert. Recursive self-organizing network models. *Neural Networks*, 17(8–9):1061–1085, 2004.
- [29] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New York, 1999.
- [30] D. O. Hebb. *The Organization of Behavior*. John Wiley & Sons Inc, New York, 1949.
- [31] H. Jaeger. Short term memory in echo state networks. Technical Report GMD Report 152, German National Research Center for Information Technology, 2001.
- [32] P. Jedlička. Synaptic plasticity, metaplasticity and BCM theory. *Bratislavské lekárske listy*, 103:137–143, 2002.
- [33] M. I. Jordan. Serial order: A parallel distributed processing approach. In *Advances in Connectionist Theory: Speech*. Erlbaum, Hillsdale, 1989.
- [34] S. Kaski, J. Kangas, and T. Kohonen. Bibliography of self-organizing maps: papers: 1981-1997. *Neural Computing Surveys*, 1:102–350, 1998.
- [35] M. Kc, M. Hagenbuchner, A. C. Tsoi, F. Scarselli, A. Sperduti, and M. Gori. XML document mining using contextual self-organizing maps for structures. In *INEX*, volume 4518 of *Lecture Notes in Computer Science*, pages 510–524, 2006.
- [36] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.
- [37] T. Kohonen. *Self-Organization and Associative Memory*. Springer, Berlin, 1984.

- [38] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.
- [39] T. Kohonen. *Symp. On Neural Networks; Alliances and Perspectives in Senri*. Senri Int. Information Institute, Osaka, Japan, 1992.
- [40] T. Kohonen. *Self-organizing maps*. Springer, Berlin, 3rd edition, 2001.
- [41] T. Kohonen and P. Somervuo. Self-organizing maps of symbol strings. *Neurocomputing*, 21(Issues 1–3):19–30, 1998.
- [42] T. Koskela, M. Varsta, J. Heikkonen, and K. Kaski. Temporal sequence processing using Recurrent SOM. In *Proceedings of the 2nd International Conference on Knowledge-Based Intelligent Engineering Systems*, pages 290–297, 1998.
- [43] T. Koskela, M. Varsta, J. Heikkonen, and K. Kaski. Time series prediction using recurrent SOM with local linear models. *International Journal of Knowledge-Based Intelligent Eng. Systems*, 2(1):60–68, 1998.
- [44] B. J. A. Kröse and M. Eecen. A self-organizing representation of sensor space for mobile robot navigation. *Proc. IROS '94*, 1:9–14, 1994.
- [45] A. Kurz. Building maps for path-planning and navigation using learning classification of external sensor data. *Artificial Neural Networks*, 1(2):587–590, 1992.
- [46] V. Kvasnička, Ľ. Beňušková, J. Pospíchal, I. Farkaš, P. Tiňo, and A. Král'. *Introduction to the Theory of Neural Networks*. IRIS, Bratislava, 1997.
- [47] M. Lukosevicius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.
- [48] T. Martinetz and K. Schulten. A neural-gas network learns topologies. In *Proceedings of the International Conference on Artificial Neural Networks*, pages 397–402, Amsterdam, 1991. North-Holland.
- [49] W. S. McCulloch and W. Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
- [50] M. Minsky and S. Papert. *Perceptrons*. MIT Press, Cambridge, Massachusetts, 1969.

- [51] R. Mori, Y. Bengio, and R. Cardin. Speaker independent speech recognition with neural networks and speech knowledge. In *Advances in neural information processing systems 2*, pages 218–225. Morgan Kaufmann Publishers Inc., 1989.
- [52] P. Návrát, M. Bieliková, Ľ. Beňušková, I. Kapustík, and M. Unger. *Umelá inteligencia*. Vydavateľstvo STU, Bratislava., 2002.
- [53] N. Neubauer. Recursive SOMs and Automata. Master’s thesis, Cognitive Science, University of Osnabrück, 2005.
- [54] C. Palmer. Sequence memory in music performance. *Current Directions in Psychological Science*, 14:247–250, 2005.
- [55] M. Pöllä, T. Honkela, and T. Kohonen. Bibliography of self-organizing map (SOM) papers: 2002-2005. Unpublished SOM bibliography, 2006.
- [56] J. Pollack. Recursive distributed representations. *Artificial Intelligence*, 46(1-2):77–105, 1990.
- [57] H. Ritter. Self-organizing maps for robot control. *International Conference on Artificial Neural Networks*, 1327 1997:673–684, 1997.
- [58] H. Ritter and T. Kohonen. Self-organizing semantic maps. *Biological Cybernetics*, 61:241–254, 1989.
- [59] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958.
- [60] D. E. Rumelhart, G. Hinton, and R. J. Williams. *Parallel Distributed Processing*. The MIT Press, Cambridge, MA, 1986.
- [61] J. Šíma and R. Neruda. *Teoretické otázky neuronových sítí*. Matfyz Press, 1997.
- [62] P. Sinčák and G. Andrejková. *Neurónové siete (Inžiniersky prístup)*, volume 1. Elfa s.r.o., 1996.
- [63] P. Sinčák and G. Andrejková. *Neurónové siete (Inžiniersky prístup)*, volume 2. Elfa s.r.o., 1996.
- [64] A. Sperduti and A. Starita. Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, 8(3):714–735, 1997.

- [65] J. Steil, R. Koiva, and A. Sperduti. Unsupervised clustering of continuous trajectories of kinematic trees with SOM-SD. In *Proceedings of the 14th European Symposium on Artificial Neural Networks*, pages 1–6, Bruges, Belgium, 2006.
- [66] M. Strickert and B. Hammer. Unsupervised recursive sequence processing. In *Neurocomputing*, pages 433–439. D-side Publications, 2003.
- [67] M. Strickert and B. Hammer. Self-organizing context learning. In *European Symposium on Artificial Neural Networks*, pages 39–44, 2004.
- [68] M. Strickert and B. Hammer. Merge SOM for temporal data. *Neurocomputing*, 64:39–71, 2005.
- [69] W. Tai. A batch training network for self-organization. In *International Conference on Artificial Neural Networks*, volume 2, pages 33–37, 1995.
- [70] P. Tiño, I. Farkaš, and J. van Mourik. Dynamics and topographic organization of recursive self-organizing maps. *Neural Computation*, 18(10):2529–2567, 2006.
- [71] University of California Irvine. Machine learning repository. internet, 2009. <http://archive.ics.uci.edu/ml/>.
- [72] P. Vančo. Dynamika dávkového učenia na modeloch samoorganizujúcich sa máp. *Kognícia a umelý život*, 9:357–362, 2009.
- [73] P. Vančo. Visualization of simple XML data using recursive self-organizing neural maps. *Informatics 2009 : International Conference on Informatics*, 10:341–346, 2009.
- [74] P. Vančo. Dekódovanie štruktúrovaných dát z natrénovaných rekurentných SOM. *Kognícia a umelý život*, 10, 2010. Submitted.
- [75] P. Vančo and I. Farkaš. Recursive self-organizing networks for processing tree structures: Empirical comparison. In *IJCCI 2009 : Proceedings of the International Joint Conference on Computational Intelligence*, volume 64, pages 459–466, 2009.
- [76] P. Vančo and I. Farkaš. Experimental comparison of recursive self-organizing maps for processing tree-structured data. *Neurocomputing*, 73(7–9):1362–1375, 2010.
- [77] T. Voegtlin. *Neural Networks and Self-Reference*. PhD thesis, Universite Lyon 2, 2002.
- [78] T. Voegtlin. Recursive self-organizing maps. *Neural Networks*, 15(8-9):979–992, 2002.

- [79] P. J. Werbos. Backpropagation through time: What does it do and how to do it. *Proceedings of IEEE*, 78:1550–1560, 1990.
- [80] R. J. Williams and D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1(2):270–280, 1989.

6 Zoznam relevantných publikovaných prác autora

- [81] I. Farkaš and P. Vančo. Spracovanie postupností symbolov pomocou rekurzívnych neurónových máp. *Kognície a umělý život*, 7:99–105, 2007.
- [82] P. Vančo. Dynamika dávkového učenia na modeloch samoorganizujúcich sa máp. *Kognícia a umelý život*, 9:357–362, 2009.
- [83] P. Vančo. Visualization of simple XML data using recursive self-organizing neural maps. *Informatics 2009 : International Conference on Informatics*, 10:341–346, 2009.
- [84] P. Vančo. Dekódovanie štruktúrovaných dát z natrénovaných rekurentných SOM. *Kognícia a umelý život*, 10, 2010. Accepted.
- [85] P. Vančo and I. Farkaš. Recursive self-organizing networks for processing tree structures: Empirical comparison. In *IJCCI 2009 : Proceedings of the International Joint Conference on Computational Intelligence*, volume 64, pages 459–466, 2009.
- [86] P. Vančo and I. Farkaš. Experimental comparison of recursive self-organizing maps for processing tree-structured data. *Neurocomputing*, 73(7–9):1362–1375, 2010.

7 Summary

The thesis deals with selected models of recursive self-organizing maps (SOMSD, MSOM and RecSOM) that have recently been extended for processing complex data types, namely tree structures and acyclic graphs. Regarding MSOM we argue that despite the commutativity operation in context computation it can distinguish between the branches of trees with permuted children, and support our theoretical claim with computer simulation.

We experimentally compare the recursive SOMs using three data sets of increasing complexity. For comparison we introduce and apply six quantitative measures focusing on different aspects of the trained maps.

Next, the practical usage of these recursive maps is experimentally shown on data sets encoded in XML format. Visualization and clusterization using these models is used for data mining. Batch learning for SOMs is presented with its advantages and disadvantages. It is shown experimentally when batch learning, in its current form, cannot be used. Modified learning algorithm is proposed. Batch learning for recursive and recurrent SOMs is introduced that enables distributed computation.

Finally, methods for data extraction constructed from the trained maps are introduced. When the complete data cannot be completely retrieved, the data reconstruction methods are proposed, either in a form of a lookup table or based on a feedforward neural network.