Probabilistische Graphische Modelle in einer Neuronalen Welt: Wie können zwei Paradigmen vereint werden?

Roman Klinger

17. Juli 2020

Die aktuelle Forschung zu maschinellem Lernen wird durch tiefe neuronale Netze dominiert. Insbesondere in der Verarbeitung von strukturierten Daten wie Bildern oder Text wurden in der Vergangenheit aber insbesondere auch probabilistische graphische Modelle untersucht. Ich vergleiche in diesem Vortrag beide Paradigmen, vergleiche Lern- und Inferenzverfahren und stelle dann an Hand von Fallstudien Möglichkeiten der Kombination vor.

Literaturverzeichnis

Überblick zu probabilistischen Graphischen Modellen (ab Folie 8):

- C. Bishop (2006). *Pattern Recognition and Machine Learning*. Springer
- D. Koller und N. Friedman (2009). *Probabilistic Graphical Models – Principles and Techniques*. MIT Press
- C. Beierle und G. Kern-Isberner (2003). *Methoden wissensbasierter Systeme*. Vieweg

Beispiele für gerichtete Modelle (Folie 11)

- Naïve Bayes: P. Domingos und M. Pazzani (1997). "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss". In: *Machine Learning* 29, S. 103–130. DOI: 10.1023/A:1007413511361
- Hidden Markov Model: L. R. Rabiner (1989). "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77.2, S. 257–286. DOI: 10.1109/5.18626
- Latent Dirichlet Allocation: D. M. Blei, A. Y. Ng und M. I. Jordan (2003). "Latent dirichlet allocation". In: *Journal of machine Learning research* 3, S. 993–1022. DOI: 10.1162/jmlr.2003.3.4-5.993

Faktorgraphen:

- F. R. Kschischang, B. J. Frey und H.-A. Loeliger (2001). "Factor graphs and the sum-product algorithm". In: *IEEE Transactions on Information Theory* 47.2, S. 498–519

- Conditional Random Fields (Folie 19): J. D. Lafferty, A. McCallum und F. C. N. Pereira (2001). "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data". In: *Proceedings of the Eighteenth International Conference on Machine Learning*. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., S. 282–289. ISBN: 1558607781. DOI: 10.5555/645530.655813

Überblick zu künstlichen neuronale Netzen (ab Folie 27:

- R. Rojas (1996). *Neural Networks – A Systematic Introduction*. Springer
- D. Patterson (1997). *Künstliche Neuronale Netze*. Prentics Hall

Details zum Vergleich von Belief Propagation und Backpropagation (Folie 38):

- J. Eisner (Nov. 2016). "Inside-Outside and Forward-Backward Algorithms Are Just Backprop (tutorial paper)". In: *Proceedings of the Workshop on Structured Prediction for NLP*. Austin, TX: Association for Computational Linguistics, S. 1–17. DOI: 10.18653/v1/W16-5901. URL: https://www.aclweb.org/anthology/W16-5901
- J. Dauwels, S. Korl und H.-A. Loeliger (2005). "Steepest Descent as Message Passing". In: *Proc. of IEEE ISOC ITW2005 on Coding and Complexity*, S. 42–46. DOI: 10.1109/ITW.2005.1531853
- J. Dauwels, S. Korl und H.-A. Loeliger (2006). "Particle Methods as Message Passing". In: *2006 IEEE International Symposium on Information Theory*, S. 2052–2056. DOI: 10.1109/ISIT.2006.261910

Restricted Boltzmann Machines (Folie 43):

- I. Goodfellow, Y. Bengio und A. Courville (2016). *Deep Learning*. MIT Press
- H. Yu (2017). *A gentle tutorial on Restricted Boltzmann Machine and Contrastive Divergence*. Online. URL: https://doi.org/10.13140/RG.2.2.26119.60326

Hidden CRF (Folie 45):

- O. Täckström und R. McDonald (Juni 2011). "Semi-supervised latent variable models for sentence-level sentiment analysis". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics, S. 569–574. URL: https://www.aclweb.org/anthology/P11-2100
- A. Quattoni, S. Wang, L.-P. Morency, M. Collins und T. Darrell (2007). "Hidden Conditional Random Fields". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.10, S. 1848–1853. DOI: 10.1109/TPAMI.2007.1124

LSTM-CRF (ab Folie 49):

- NER (Folie 50): Z. Huang, W. Xu und K. Yu (2015). *Bidirectional LSTM-CRF Models for Sequence Tagging*. URL: http://arxiv.org/abs/1508.01991
- Bildsegmentierung (Folie 51): S. Zheng, S. Jayasumana u. a. (2015). "Conditional Random Fields as Recurrent Neural Networks". In: *International Conference on Computer Vision (ICCV)*. URL: https://www.robots.ox.ac.uk/~szheng/papers/CRFasRNN.pdf
- Autonomes Fahren (Folie 52): X. Wang, J. Wu, Y. Gu, H. Sun, L. Xu, S. Kamijo und N. Zheng (2018). "Human-Like Maneuver Decision Using LSTM-CRF Model for On-Road Self-Driving". In: *21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui, Hawaii, USA. DOI: 10.1109/ITSC.2018.8569524

In Relation stehende Sequenzen:

- Factorial Conditional Random Fields (Folie 53): C. Sutton, A. McCallum und K. Rohanimanesh (2007). "Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data". In: *Journal of Machine Learning Research* 8.Mar, S. 693–723
- Neural Factor Graph Models (Folie 54): C. Malaviya, M. R. Gormley und G. Neubig (Juli 2018). "Neural Factor Graph Models for Cross-lingual Morphological Tagging". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, S. 2653–2663. DOI: 10.18653/v1/P18-1247. URL: https://www.aclweb.org/anthology/P18-1247

Relationserkennung in Text (Folie 56):

- M. R. Gormley, M. Yu und M. Dredze (Sep. 2015). "Improved Relation Extraction with Feature-Rich Compositional Embedding Models". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, S. 1774–1784. DOI: 10.18653/v1/D15-1205. URL: https://www.aclweb.org/anthology/D15-1205
- N. FitzGerald, O. Täckström, K. Ganchev und D. Das (Sep. 2015). "Semantic Role Labeling with Neural Network Factors". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, S. 960–970. DOI: 10.18653/v1/D15-1112. URL: https://www.aclweb.org/anthology/D15-1112
- C. Lyu und I. Titov (Juli 2018). "AMR Parsing as Graph Prediction with Latent Alignment". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, S. 397–407. DOI: 10.18653/v1/P18-1037. URL: https://www.aclweb.org/anthology/P18-1037

Integrierte Werkzeuge (Folie 58)

- PyTorch CRF-Schicht, https://pytorch-crf.readthedocs.io/en/stable/
- AllenNLP CRF-Schicht, https://github.com/allenai/allennlp/blob/master/allennlp/modules/conditional_random_field.py
- Torch-Struct, https://github.com/harvardnlp/pytorch-struct, A. Rush (Juli 2020). "Torch-Struct: Deep Structured Prediction Library". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Online: Association for Computational Linguistics, S. 335–342. URL: https://www.aclweb.org/anthology/2020.acl-demos.38
- SPEN, https://github.com/davidBelanger/SPEN, D. Belanger und A. McCallum (20–22 Jun 2016). "Structured Prediction Energy Networks". In: *Proceedings of The 33rd International Conference on Machine Learning*. Hrsg. von M. F. Balcan und K. Q. Weinberger. Bd. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, S. 983–992. URL: http://proceedings.mlr.press/v48/belanger16.html
- lp-sparsemap, https://github.com/deep-spin/lp-sparsemap, V. Niculae und A. F. T. Martins (2020). *LP-SparseMAP: Differentiable Relaxed Optimization for Sparse Structured Prediction*. arXiv: 2001.04437 [cs.LG]. URL: https://arxiv.org/abs/2001.04437

Software (Folie 62):

- FACTORIE, http://factorie.cs.umass.edu/, A. McCallum, K. Schultz und S. Singh (2009). "FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs". In: *Neural Information Processing Systems (NIPS)*
- Markov-Logik: M. Richardson und P. Domingos (2006). "Markov logic networks". In: *Machine Learning* 62.1–2, S. 107–136. DOI: 10.1007/s10994-006-5833-1
- Alchemy: http://alchemy.cs.washington.edu/
- Markov the Beast, https://code.google.com/archive/p/thebeast/, S. Riedel (2008). "Improving the accuracy and Efficiency of MAP Inference for Markov Logic". In: *Proceedings of the 24th Annual Conference on Uncertainty in AI (UAI)*, S. 468–475
- Infer.Net/CSoft: https://dotnet.github.io/infer/, S. S. J. Wang und M. P. Wand (2011). "Using Infer.NET for Statistical Analyses". In: *The American Statistician* 65.2, S. 115–126. DOI: 10.1198/tast.2011.10169

Weitere relevante Arbeiten, die ich nicht besprochen habe:

- M. J. Johnson, D. K. Duvenaud, A. Wiltschko, R. P. Adams und S. R. Datta (2016). "Composing graphical models with neural networks for structured representations and fast inference". In: *Advances in Neural Information Processing Systems 29*. Hrsg. von D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon und R. Garnett. Curran Associates, Inc., S. 2946–2954. URL: http://papers.nips.cc/paper/6379-composing-graphical-models-with-neural-networks-for-structured-representations-and-fast-inference.pdf
- M. R. Gormley (2015). "Graphical Models with structured factors, neural factors, and approximation-aware training". Diss. Baltimore, USA: Johns Hopkins University. URL: http://jhir.library.jhu.edu/handle/1774.2/39628
- Z. Zhang, F. Wu und W. S. Lee (2019). *Factor Graph Neural Network*. arXiv: 1906.00554 [cs.LG]. URL: https://arxiv.org/abs/1906.00554
- G. Emerson und A. Copestake (Aug. 2016). "Functional Distributional Semantics". In: *Proceedings of the 1st Workshop on Representation Learning for NLP*. Berlin, Germany: Association for Computational Linguistics, S. 40–52. DOI: 10.18653/v1/W16-1605. URL: https://www.aclweb.org/anthology/W16-1605