prominent in the case of the proposed models, compared to the conventional CRBM formulations. Finally, we also note the notable performance gains of all the considered energy-based models compared to the PYP-GP method, which relies on a much more computationally expensive Gaussian process model to perform sequential data modeling.

Conclusions

In this paper, we proposed a method exploiting the merits of ESNs to enhance the sequential data modeling capabilities of CRBMs. Our approach consists in the utilization of an ESN reservoir to capture the temporal dynamics in the context of CRBMs instead of the linear autoregressive apparatus of existing approaches. This model formulation allows for extracting more complex temporal dynamics using less *trainable* model parameters.

Subsequently, we extended the so-obtained ES-CRBM model to obtain an implicit mixture of ES-CRBM experts, capable of better modeling multimodal sequential data, as well as performing classification of observed sequences, apart from data modeling and generation. Exact inference for our models was performed by means of an elegant alternating Gibbs sampling algorithm, while training was conducted by means of CD.

Our future goals focus on investigating the efficacy of our approach in a wide class of applications involving time-series data or data with temporal dynamics. Characteristic application areas include dynamic planning algorithms for multirobot swarms, automatic music improvisation and metacreation, and analysis and prediction of asset prices in financial markets using high-frequency measurements.

References

Altun, Y.; Tsochantaridis, I.; and Hofmann, T. 2004. Hidden Markov support vector machines. In *Proc. ICML*.

Bakker, B., and Heskes, T. 2003. Task clustering and gating for Bayesian multitask learning. *Journal of Machine Learning Research* 4:83–99.

Bengio, Y., and Delalleau, O. 2008. Justifying and generalizing contrastive divergence. *Neural Computation* 21(1):1–21.

Chatzis, S. P., and Demiris, Y. 2012. Nonparametric mixtures of Gaussian processes with power-law behavior. *IEEE Transactions on Neural Networks and Learning Systems* 23(12):1862–1871.

Haykin, S., and Principe, J. 1998. Making sense of a complex world. *IEEE Signal Process. Mag.* 15(3):66–81.

Hinton, G. E., and Salakhutdinov, R. 2006. Reducing the dimensionality of data with neural networks. *Science* 313(5786):504 – 507.

Hinton, G. E. 2002. Training products of experts by minimizing contrastive divergence. *Neural Computation* 14(8):1771–1800.

Kianifardand, F., and Swallow, W. 1996. A review of the development and application of recursive residuals in linear models. *J. Amer. Statist. Assoc.* 91(443):391–400.

Kosmopoulos, D., and Chatzis, S. 2010. Robust visual behavior recognition. *Signal Processing Magazine, IEEE* 27(5):34 –45.

Larochelle, H.; Erhan, D.; Courville, A.; Bergstra, J.; and Bengio, Y. 2007. An empirical evaluation of deep architectures on problems with many factors of variation. In *Proc. ICML*, 473–480. ACM Press.

Lukosevicius, M., and Jaeger, H. 2009. Reservoir computing approaches to recurrent neural network training. *Computer Science Review* 3:127–149.

Maass, W.; Natschlaeger, T.; and Markram, H. 2002. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation* 14(11):2531–2560.

McLachlan, G., and Peel, D. 2000. *Finite Mixture Models*. New York: Wiley Series in Probability and Statistics.

Nair, V., and Hinton, G. 2008. Implicit mixtures of restricted Boltzmann machines. In *Proc. NIPS*.

Ni, B.; Wang, G.; and Moulin, P. 2011. RGBD-HuDaAct: A color-depth video database for human daily activity recognition. In *ICCV Workshops*, 1147–1153.

Sha, F., and Saul, L. K. 2007. Comparison of large margin training to other discriminative methods for phonetic recognition by hidden Markov models. In *Proc. ICASSP*, 313–316.

Smolensky, P. 1986. Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, e. a., ed., *Parallel Distributed Processing: Volume 1: Foundations*, 194–281. Cambridge, MA: MIT Press.

Taylor, G. W.; Sigal, L.; Fleet, D. J.; and Hinton, G. E. 2010. Dynamical binary latent variable models for 3d human pose tracking. In *Proc. CVPR*, 631–638.

Taylor, G. W.; Hinton, G. E.; and Roweis, S. T. 2011. Two distributed-state models for generating high-dimensional time series. *J. Machine Learning Research* 12:1025–1068.

Verstraeten, D.; Schrauwen, B.; D'Haene, M.; and Stroobandt, D. 2007. 2007 special issue: An experimental unification of reservoir computing methods. *Neural Networks* 20(3):391–403.

Ye, N.; Lee, W. S.; Chieu, H. L.; and Wu, D. 2009. Conditional random fields with high-order features for sequence labeling. In *Proc. NIPS*.

Zhu, J.; Chen, N.; and Xing, E. P. 2011. Infinite SVM: a Dirichlet process mixture of large-margin kernel machines. In *Proc. ICML*.