HIDDEN MARKOV MODELS IN CONTEXT

JOHN ASHLEY BURGOYNE · ID 260172518

MUMT 611 · PROF. ICHIRO FUJINAGA · 1 FEBRUARY 2007

HIDDEN MARKOV MODELS (HMMS), following on their great success in speech recognition (Rabiner 1989), have been used with great success in music information retrieval (MIR), and for good reason (Raphael 1999; Sheh and Ellis 2003). As machine-learning models go, they are intuitive, and good implementations of the basic algorithms for training and applying HMMS are freely available for many common programming languages. They handle time-ordered sequences naturally, which is important for many applications in the field. Nonetheless, they are not always an optimal or even appropriate choice for certain tasks. In these cases, it worth exploring alternatives to the HMM, and as yet, few researchers in MIR have done so.

GENERATIVE MODELS

Hidden Markov models are part of a family known as generative models, probablisitic models that, once trained, are capable of generating new data given some initial condition. The most basic generative model is a Gaussian distribution, which would always generate its mean regardless of initial conditions. By various transformations of the Gaussian distribution, such as mixing multiple models, reducing dimensionality, adding time dynamics to the system, or adding hierarchy, one can derive a variety of other common statistical models, including principal component analysis, independent component analysis, vector quantisation, and several variants of the HMM (Murphy 1998; Roweis and Ghahramani 1999).

The first set of alternative models to consider are the basic variants of the HMM, including the autoregressive HMM, which allows for time dependency among observations as well as states, and the input-output HMM, which allows each state to take an input as well as an output. Another tempting set of models to consider is that of grouped HMMs, such as the factorial HMM, which comprises a group of parallel Markov chains sharing common outputs, and the coupled HMM, in which several time-synchronised HMMs attempt to share information across their states. These grouped models seem to describe many musical phenomena well, e.g., polyphony, but unfortunately, they are so difficult to train that they tend to be unsuccessful in practice (Murphy 2002). The Kalman filter and its derivatives, best known in musical applicaitons for adaptive IIR filtering, are the models of choice to use when an HMMwould otherwise work except for the fact that its states would need to be continuous rather than discrete (Welch and Bishop 1995).

DISCRIMINATIVE MODELS

Discriminative models, or probabilistic models in which the states are dependent on the observed data, can be more flexible than generative ones. Traditional discriminative models are context-independent classifiers, but more recently, discriminative models have been developed that can handle sequences. The advantage of using a discriminative model for a sequence rather than a generative one is that a discriminative model considers the complete sequence simultaneously, rather than enforcing the Markov assumption of an HMM. In their simplest form, however, these models differ only slightly from the HMM: by reversing the direction of the arrows between states an observation to make each state of an HMMdependent on its corresponding observation, one generates a maxmimum entropy Markov model (MEMM). These models have proven to be remarkably successful in sequential classification tasks (McCallum, Freitag, and Pereira 2000).

The MEMM has been developed further into a more general model based on Markov random fields (MRFs) rather than Markov chains. This model is known as the conditional random field (CRF) has been even more successful than the MEMM for sequential classification tasks (Lafferty, McCallum, and Pereira 2001). Their principal disadvantage is their complexity of implementation and heavy computational demands (Sutton and McCallum 2006). Code libraries for CRFs do exist, however, and as these models mature, we can expect the number of such libraries to increase.

Perhaps because of the difficulties in implementaiton, no author in the MIR community has used these approaches to date, although there are moves in this direction. Poliner and Ellis (forthcoming) have used a simpler discriminative model for piano transcription that breaks the time dependency, which requires them to use an HMMto smooth the output. Pickens and Iliopoulos (2005) have used MRFs for modeling music, but in the generative form, not the discriminative CRF described here. It appears to be a fruitful direction for future work.

REFERENCES

Lafferty, J., A. McCallum, and F. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, Williamstown, Mass., pp. 282–89.

McCallum, A., D. Freitag, and F. Pereira. 2000. Maximum entropy Markov models for information extraction and segmentation. In *Proceedings of the 17th International Conference on Machine Learning*, Stanford, Calif., pp. 591–98.

Murphy, K. 1998. A brief introduction to graphical models and Bayesian networks. Online tutorial. http://www.cs.ubc.ca/ murphyk/Bayes/bnintro.html.

Murphy, K. 2002. A tutorial on dynamic Bayesian networks.

Presentation to the MIT AI lab. http://www.cs.ubc.ca/ murphyk/Papers/dbntalk.pdf.

Pickens, J., and C. Iliopoulos. 2005. Markov random fields and maximum entropy modeling for music information retrieval. In *Proceedings of the 6th International Conference on Music Information Retrieval*, London, U.K., pp. 207–14.

Poliner, G. E., and D. P. W. Ellis. Forthcoming. A discriminative model for polyphonic piano transcription. *EURASIP Journal on Advances in Signal Processing*. [n.p.]

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77 (2): 257–87.

Raphael, C. 1999. Automatic segmentation of acoustic musical signals using hidden Markov models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (4): 360–70.

Raphael, C. 2001. A probabilistic expert system for automatic musical accompaniment. *Journal of Computational and Graphical Statistics* 10 (3): 487–512.

Roweis, S., and Z. Ghahramani. 1999. A unifying review of linear Gaussian models. *Neural Computation* 11 (2): 205–45.

Sheh, A., and D. P. W. Ellis. 2003. Chord segmentation and recognition using EM-trained hidden Markov models. In *Proceedings of 4th International Conference on Music Information Retrieval*, Baltimore, Md., pp. 185–91.

Sutton, C., and A. McCallum. 2006. An introduction to conditional random fields for relational learning. In L. Getoor and B. Taskar (Eds.), *Introduction to Statistical Relational Learning*, [n.p.]. MIT Press.

Welch, G., and G. Bishop. 1995. An introduction to the Kalman filter. Technical Report 95-041, Univ. of N.C. at Chapel Hill, Dept. of Computer Science.