Self-organizing mixture models

J.J. Verbeek N. Vlassis B.J.A. Kröse

Informatics Institute, University of Amsterdam

The full version of this paper will appear as [4].

Summary

We present an expectation-maximization (EM) algorithm that yields topology preserving maps of data based on probabilistic mixture models. Our approach is applicable to any mixture model for which we have a normal EM algorithm. Compared to other mixture model approaches to self-organizing maps, the function our algorithm maximizes has a clear interpretation: it sums data log-likelihood and a penalty term that enforces self-organization. Our approach allows principled handling of missing data and learning of mixtures of self-organizing maps. We present example applications illustrating our approach for continuous, discrete, and mixed discrete and continuous data.

The self-organizing map, or SOM for short, was introduced by Kohonen in the early 1980's; it combines clustering of data with topology preservation. The clusters found in the data are represented on a, typically two dimensional, grid, such that clusters with similar content are nearby in the grid. The representation thus preserves topology in the sense that it keeps similar cluster nearby. The SOM allows one to visualize high dimensional data in two dimensions, e.g. on a computer screen, via a projection that may be a non-linear function of the original features in which the data was given. The clusters are sometimes also referred to as 'nodes', 'neurons' or 'prototypes'. Since their introduction, self-organizing maps have been applied in many engineering problems, see e.g. the list of over 3000 references on applications of SOM in [2].

Probabilistic mixture models are densities that can be written as a weighted sum of component densities, where the weighting factors are all non-negative and sum to one [3]. By using weighted combinations of simple component densities a rich class of densities is obtained. Mixture models are used in a wide range of settings in machine-learning and pattern recognition. Examples include classification, regression, clustering, data visualization, dimension reduction, etc. A mixture model can be interpreted as a model that assumes that there are, say k, sources that generate the data: each source is selected to generate data with a probability equal to its mixing weight and it generates data according to its component density. Marginalizing over the components, we recover the mixture model as the distribution over the data. With a mixture model we can associate a clustering of the data, by assigning each data item to the source that is most likely to have generated the data item. The expectation-maximization (EM) algorithm [1] is a popular algorithm to fit the parameters of a mixture to given data.

Several variations of the original SOM algorithm have been proposed in the literature, they can be roughly divided into two groups. First, different distance measures have been proposed to assign data points to clusters when using different types of data. Second, alternative learning algorithms for SOMs have been proposed. In this paper we show how to combine the benefits of self-organizing maps and mixture models. We present a general learning algorithm, similar to Kohonen's original SOM algorithm.

In contrast to Kohonen's original algorithm, ours is guaranteed to converge and can be interpreted as maximizing an objective function that sum the data log-likelihood and a penalty term. The algorithm is the standard EM learning algorithm using a slightly modified expectation-step. Our contribution can be considered as one in the category of SOM papers presenting new learning algorithms. Since we merely modify the expectation-step, we can directly make a self-organizing map version of any mixture model for which we have a normal EM algorithm. We only need to replace normal E-step with the modified E-step presented here to obtain a self-organizing map version of the given mixture model. Thus, because our modified EM algorithm can be applied to any mixture model for which we have a normal EM algorithm, it can be applied to a wide range of data types. Prior knowledge or assumptions about the data can be reflected by choosing an appropriate mixture model. The mixture model will therefore, implicitly, provide a suitable distance measure. We can thus state that our work also gives a contribution of the first type: it helps to design distance measures implicitly by specifying a generative model for the data. Like many other probabilistic models, our approach also allows for principled handling of missing data values and offers the ability to learn mixtures of self-organizing maps.

In our paper [4] we first review self-organizing maps and the EM algorithm for mixture models and motivate our work. After the presentation of our mixture model based approach to self-organizing maps, we compare it to related existing work. We also present experimental results obtained by applying our method to several example data sets with discrete variables, continuous variables and a data set which has both.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[2] T. Kohonen. *Self-organizing maps.* Springer, 2001.

[3] G. J. McLachlan and D. Peel. *Finite mixture models.* Wiley, New-York, NY, USA, 2000.

[4] J. J. Verbeek, N. Vlassis, and B. J. A. Kröse. Self-organizing mixture models. *Neurocomputing*, 2004. to appear.