Data Driven Generation of Interactions for Feature Binding and Relaxation Labeling

Sebastian Weng and Jochen J. Steil

Neuroinformatics Department, P.O.-Box 100131, D-33501 Bielefeld, Germany, {sweng, jsteil}@techfak.uni-bielefeld.de,

Abstract. We present a combination of unsupervised and supervised learning to generate a compatibility interaction for feature binding and labeling problems. We focus on the unsupervised data driven generation of prototypic basis interactions by clustering of proximity vectors computed from pairs of data in the training set. Subsequently a supervised method recently introduced by [9] is used to determine coefficients to form a linear combination of the basis functions, which then serves as interaction. As special labeling dynamic we use the competitive layer model, a recurrent neural network with linear threshold neurons, and shown an application to cell segmentation.

1 Introduction

Feature binding and relaxation labeling have widely been used to solve many segmentation tasks occurring in early vision, perceptual grouping, and pattern matching. To bind features to groups, a priori knowledge on the degree of compatibility of pairs of features is encoded in an interaction function, which in most cases is derived heuristically and fixed beforehand. Successful application are reported in the areas of point clustering, contour grouping [10], texture segmentation [3], and cell segmentation [4]. However, in more complex and more abstract feature spaces a heuristic and suitable encoding of contextual knowledge in the interaction becomes increasingly difficult and motivates the search for learning methods.

One of the main bottlenecks is, that when deriving the interaction directly from a low number of training patterns the number of feature pairs may be large, but nevertheless spans only a small subspace of all possible pairwise interactions. Several approaches to learning the compatibility function have been proposed [5,6,7], but lead to very high dimensional nonlinear optimization problems, which suffer from local minima and respective poor generalization. In Wersing [9] this problem is solved by reduction to the task of finding weighting coefficients for a linear combination of a number of interaction prototypes, however at the cost of heuristically choosing the basis functions. In this contribution we show, how appropriate basis functions can be generated using unsupervised learning.

Starting point for the formal derivation of the learning problem is the concept of consistency [8] to derive a set of linear inequalities describing the learning
problem. To segment a pattern consisting of feature vectors $m_r^i$ a label $\alpha = 1, \ldots, L$ has to be assigned to each $m_r$, employing pairwise interactions $f_{\alpha\beta} = f(m_r, m_r^\beta)$. Positive values of $f_{\alpha\beta}$ indicate that $m_r$ and $m_r^\beta$ have a tendency to get the same label, while negative give preference to separate them. Let $0 \leq x_{r\alpha} \leq 1$ represent the grade of assignment feature $m_r$ to the label $\alpha$, then a consistent labeling implies, that the mutual support a feature receives from the assignment to the label $\bar{\alpha}$ is larger than in case of using a different label $\beta$:

$$\sum_{r'} f_{\alpha r'} x_{r'\beta} < \sum_{r'} f_{\alpha r'} x_{r'\bar{\alpha}} \quad \text{for all } r, \beta \neq \bar{\alpha}(r). \quad (1)$$

This consistency condition (1) has been shown to imply that the corresponding labeling is a asymptotically stable attractor of the common relaxation labeling dynamics [8]. However, instead of iterating a classical relaxation dynamics in this contribution we employ the competitive layer model (CLM), a large-scale, topographically ordered, linear threshold network, because it’s stable states lead to the same consistency conditions [10]. This is especially attractive, because recently developed efficient hardware implementations [1] are available.

In Section 2 we formulate the learning problem based on the consistency approach (1) and introduce the decomposition of $f_{\alpha\beta}$ in basis prototypes. In Section 3 we employ a self organizing map for data driven generation of suitable prototypes. In Section 4 we apply this approach to a figure-background segmentation of cell images and give a discussion in the final Section 5.

2 The Learning Problem

The goal is to generate the desired interactions $f_{\alpha\beta}$ from a set of $M$ labeled training patterns $\mathcal{P}^i$, $i = 1, \ldots, M$, where each pattern consists of a subset $\mathcal{R}^i = \{r^i_1, \ldots, r^i_{N_i}\}$ of $N^i$ different features and their corresponding labels $\bar{\alpha}^i(r_j)$. For each labeled training pattern a target vector $y^i$ is constructed by choosing

$$y^i_{r\bar{\alpha}(r)} = 1, \quad y^i_{r\beta} = 0, \quad \text{for all } r \in \mathcal{R}^i, \beta \neq \bar{\alpha}(r). \quad (2)$$

where columns for features, which are not contained in the training pattern are filled with zeros according to $y^i_{rs} = 0$ for all $\alpha$, $s \not\in \mathcal{R}^i$. We obtain the following set of $(L - 1) \sum_i N^i$ consistency inequalities:

$$\sum_{r'} f_{\alpha r'} y^i_{r'\beta} < \sum_{r'} f_{\alpha r'} y^i_{r'\bar{\alpha}} \quad \text{for all } i, r \in \mathcal{R}^i, \beta \neq \bar{\alpha}. \quad (3)$$

Fixing a number of $K$ basis interactions $g^j_{\alpha\beta} = g^j(m_r, m_r^\beta)$ we represent $f_{\alpha\beta}$ as linear combination $f_{\alpha\beta} = \sum_{j=1}^{K} c_j g^j_{\alpha\beta}$, $j = 1, \ldots, K$ with weight coefficients $c_j$. Inserting into (3) we obtain the reduced problem in $K$ new parameters $c_j$ as:

$$\sum_{j} c_j \sum_{r'} g^j_{\alpha r'} (y^i_{r'\beta} - y^i_{r'\bar{\alpha}}) + \kappa < 0, \quad \text{for all } i, r \in \mathcal{R}^i, \beta \neq \bar{\alpha}, \quad (4)$$

$m_r^i$ are elements of a problem specific domain $\mathcal{F}$, e.g. the set of all possible local edge elements $(x_r, y_r, \varphi_r)$ in an image.
where an additional a positive margin variable $\kappa > 0$ was added to achieve a better robustness. Inconsistent training data or higher values of $\kappa$ imply that not all inequalities can be fulfilled and thus in Wersing [9] a quadratic approach to minimize their overall violation was utilized.

In addition to the standard relaxation labeling approach we use special label with index $\alpha = 1$ to collect all features, which have only a weak binding to the main groups and can be considered noise. To this aim the inequalities

$$\begin{align*}
\sum_j c_j \sum_{r, \alpha} g^{ij}_{r, \alpha} y^{ji}_{r, \alpha} - m + \kappa &< 0 \quad \text{for all } i, \beta, r \in R^i, \bar{\alpha}(r) = 1, \\
m - \sum_j c_j \sum_{r, \alpha} g^{ij}_{r, \alpha} y^{ji}_{r, \alpha} + \kappa &< 0
\end{align*}$$

introduce a threshold $m$. In the case a feature $r$ is labeled noisy the first $L$ inequalities in (5) express that the mutual support it receives in the group $\beta$ is below the threshold. In case that feature $r$ is assigned a label $\alpha > 1$ the one remaining inequality in (5) used the desired labels $y^{ji}_{r, \alpha}$ from the training set to enforce that the mutual support in the training data is above the threshold $m$. Once the prototypes $g^i(m_r, m_r)$ are chosen the overall learning problem consists of $L \sum_i N^i$ linear inequalities in (4) and (5) with $K + 1$ free parameters.

3 Unsupervised Learning of Interaction Prototypes

We turn now to the problem to generate suitable symmetric interaction prototypes $g^i_{r, \beta} = g^i(m_r, m_r)$. As $f_{r, \beta}$ and thus as well $g_{r, \beta}$ expresses some degree of mutual compatibility or proximity of a feature pair $m_r, m_r$ we first transform it into a generalized proximity space $D$ by a vector function

$$d_{r, \beta} = [d_1(m_r, m_r), \ldots, d_P(m_r, m_r)]^T. \quad (6)$$

Here $d_i(\cdot), i \geq 1$ may be an arbitrary norm, but as well more general problem specific compatibility function taking also negative values. Then we specify $g_{r, \beta} = g(d_{r, \beta})$ as function with arguments in $D$ (and thus implicitly $f_{r, \beta}$ as well). The elementary example is to directly use the euclidian distance $f_{r, \beta} = g_{r, \beta} = d_{r, \beta} = \| (x, y) - (x, y) \|$ between image coordinates $m_r = (x, y)$ for point clustering.

The most direct approach to specify a set of $g^i_{r, \beta}$ is to decompose the proximity vector space $D$ into disjunct regions $D^i \cup D^\beta = D$. In Wersing [9] this was done by division of each dimension $d^i_{r, \beta}$ in a set of heuristically specified disjunct intervals. But in case of more complex features and corresponding proximity data there is need for data driven learning methods.

We use a variation of self-organizing map, the activity equilibration AEV [2], to reduce the proximity vectors $d_{r, \beta}$ to a set of $K$ proximity prototypes $\bar{d}_j$ and choose as $j-th$ basis function $g^j$ the corresponding $j-th$ multidimensional Voronoi cell

$$g^j_{r, \beta} = \begin{cases} 1 & : \| d_{r, \beta} - \bar{d}_j \| \leq \| d_{r, \beta} - \bar{d}_i \| \text{ for all } i \neq j, i, j \in \{1, \ldots, K\} \\ 0 & : \text{else} \end{cases} \quad (7)$$
Figure 1 illustrates the test case of features \( m_r = (x_r, y_r, I_r) \) consisting of the two dimensional image position \((x_r, y_r)\) and a second property \(I_r\) that is displayed in form of a gray value. A two-dimensional proximity vector is formed as \( d_{r'} = [d_p, d_1]^T \), where \( d_p = \| (x_r, y_r) - (x_{r'}, y_{r'}) \| \) and \( d_1 = |I_r - I_{r'}| \).

## 4 Application

In a more complex domain, we apply the learning approach to the problem of segmentation of fluorescence cell images. The training set of ten images each showing one cell is displayed at the top Fig. 3. The desired output is a figure background segmentation of the cell body from its surrounding area. Thus the interaction function has to bind features within the cell body together and to adjust the threshold \( m \), such that the surrounding is treated as noise.

The segmentation is based on local edge feature

\[
\mathbf{m}_r = (x_r, y_r, \varphi_r), \quad x_r, y_r \in [1, 30], \quad \varphi_r \in [0, 2\pi],
\]

where \( x_r \) and \( y_r \) describe the position of a image point and \( \varphi_r \) describes the direction of the intensity gradient at this position. The construction of the proximity vectors \( \mathbf{d}_{r'} = [\| \mathbf{d} \|, \theta_1, \theta_2]^T \) is shown in Fig. 2 and follows [4]. It uses the euclidian distance between two features and angles between their connecting vector \( \mathbf{d} \) and the gradients vectors of the two features.

For the unsupervised clustering we use 30 SOM-nodes to restrict the following optimization problem to the corresponding 30 coefficients plus the threshold \( m \) for the background. The assignment inequalities are computed according to (4) and (5) and the interaction obtained for \( \kappa = 100 \) is shown in Fig. 2. The result shows remarkable generalization, the segmentation of 10 out of 30 test patterns in displayed in Fig. 3.
Fig. 2. Left: The feature vectors \( m_r(x_r, y_r, \varphi_r) \) contain position and orientation of the local grey value gradient. The proximity vectors contain \( d_r = (\| d \|, \theta_1, \theta_2) \) use the point position distance and relative angles \( \theta_1, \theta_2 \). Right: Learned interaction: For relative feature positions on a five pixel wide lattice, the strength of the interaction with the fixed feature \( m_r = (0, 0, 0) \) is visualized as grey level. The threshold \( m \) was set to 10 by the optimization algorithm.

5 Discussion

We present an approach to decompose the problem of learning a suitable compatibility interaction for feature binding and labeling problems into a new data driven step of learning prototypic basis functions, which is followed by an optimization step earlier introduced in [9]. This method resolves a heuristic step in the definition of the compatibility measure and has several advantages, for instance that the number of parameters in the optimization step is restricted by the number of SOM-prototypes used. For the case of complex feature it is even more important, that an arbitrary number of different compatibility measures can be used when transforming the data into the proximity space, because it may not be known beforehand, which are the relevant dimensions. This is clearly superior to the manual choice of a holistic proximity measure or manual tuning of coefficients or basis functions.

There are still a number of open issues, for instance more sophisticated methods like kernel or linear inequality optimization may obtain better coefficients from the consistency equations. Also could Voronoi cell based basis functions could be replaced by radial basis functions or more non-local approximations of the data in the proximity space. However, such approaches generate more interdependencies between the choice of the basis functions and the optimization of their combination, which renders the overall learning process less tractable and controllable. Nevertheless may future work provide results in this direction.
Fig. 3. First row: ten training images. Second row: the target segmentation into body and background. Third row: ten typical test images. Last row: Superimposition of target answer (light color) and CLM-segmentation with learned interaction (dark color) results in gray color for correct labels. The average proportions over 90 test images are 71% for correct labels and 9% and 20% for white and dark.

References