# Contextual Recognition of Hand-drawn Diagrams with Conditional Random Fields

Martin Szummer, Yuan Qi*
*Microsoft Research, 7 J J Thomson Avenue, Cambridge CB3 0FB, UK*
*szummer@microsoft.com, yuanqi@media.mit.edu*

## Abstract

*Hand-drawn diagrams present a complex recognition problem. Fragments of the drawing are often individually ambiguous, and require context to be interpreted.*

*We present a recognizer based on conditional random fields (CRFs) that jointly analyze all drawing fragments in order to incorporate contextual cues. The classification of each fragment influences the classification of its neighbors. CRFs allow flexible and correlated features, and take temporal information into account. Training is done via conditional MAP estimation that is guaranteed to reach the global optimum. During recognition we propagate information globally to find the joint MAP or maximum marginal solution for each fragment. We demonstrate the framework on a container versus connector recognition task.*

## 1. Introduction

Hand-drawn diagrams consist of parts whose function depends heavily on context. For example, a single line fragment could constitute the side of a container, or the stem of a connector, and its role could only be disambiguated by looking at neighboring fragments. In this paper we cleanly incorporate context into recognition of parts by treating the problem as a joint classification task. In other words, instead of recognizing parts individually, we will recognize them together by making interdependent classifications.

The recognition of parts in complex scenes is tightly intertwined with the problem of segmentation. Good segmentation is often crucial for achieving accurate recognition. Traditionally these two problems have been treated separately, by first segmenting a drawing or image into individual shapes or objects, and then recognizing each segment. The problem is that it is difficult to segment the scene into objects before recognizing them. We postpone the segmentation task and initially only divide the input into small generic fragments of objects (short stroke fragments). We then perform joint classification to identify what object classes all the fragments are parts of. This step results in an implicit segmentation, except that fragments in the same class must still be grouped, which is typically a much easier task. Thus, our joint classifier combines recognition with partial segmentation.

Combined segmentation and recognition has been done previously using generative probabilistic models. Coughlan and Ferreira [1] found objects in images using deformable templates which were fit with dynamic programming and loopy belief propagation. However, the technique depends heavily on templates, which limits its applicability. For complicated or free-form objects such as connectors, it is difficult to construct templates or generative models for the observed drawing $P(\mathbf{x})$. Recently, Tu et al. [6] proposed a Markov Chain Monte Carlo approach for image parsing, which combines segmentation, detection, and recognition. They also used a generative approach to model objects.

Unlike the above generative approaches, we advocate a discriminative classification scheme, where we only model the conditional distribution of the labels $\mathbf{y}$, i.e., $P(\mathbf{y}|\mathbf{x})$. This conditional distribution is what we ultimately need for the classification task. We avoid modeling the drawing data distribution $P(\mathbf{x})$, which can be complicated and is not needed.

Many traditional discriminative classification approaches such as logistic regression, neural networks and support vector machines can only model labeled points independently of one another. In contrast, conditional random fields (CRFs) [4] model dependencies not only between input data and its labels, but also model dependencies between labels. Specifically, they provide a joint distribution over the labels conditioned on the data. They are an instance of undirected graphical models [2].

Kumar and Hebert [3] applied conditional random fields and achieved good results on an image region classification problem. They employed a crude pseudo-likelihood approx-

* Now at MIT Media Lab, 20 Ames Street, Cambridge, MA, 02139, USA

Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition (IWFHR-9 2004)
0-7695-2187-8/04 $20.00 © 2004 IEEE

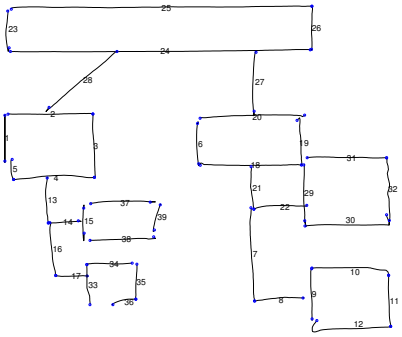

**Figure 1. An organization chart. Stroke fragments are numbered and their endpoints marked.**

imation for maximum-likelihood learning of model parameters, which tends to overestimate parameter values. They used other approximations (iterated conditional modes) for inference.

This paper describes how to apply conditional random fields to two-dimensional recognition problems exactly, without resorting to the pseudo-likelihood or other approximations. Conditional random fields show great promise in classifying sequence data (such as text or speech), and we now demonstrate their power in a two-dimensional setting, without the previously crippling assumptions.

Our application is recognition of hand-drawn organization charts (Figure 1). We focus on classifying which fragments of ink are parts of containers versus parts of connectors. This is an inherently ambiguous problem, as both containers and connectors consist of similar stroke fragments, and only the context of a fragment can resolve its true class.

In the following sections, we first introduce conditional random fields, and then derive training and inference procedures. We then describe the ink parsing task and experimental results.

## 2. Conditional Random Fields

A conditional random field can be seen as a network of interacting classifiers: the decision made by one classifier influences the decisions of its neighbors. We thus need to describe the individual classifiers, their inputs and outputs, as well as the structure of the network.

### 2.1. Formal Definition

Let $\mathbf{x}$ be an input random vector for the observed data, and $\mathbf{y}$ be an output random vector over labels of the corresponding data. The input $\mathbf{x}$ might range over the ink and the output $\mathbf{y}$ range over the labels of shapes to be recognized. All components $y_i$ of $\mathbf{y}$ are assumed to range over a set of labels $\mathcal{T}$. In this paper, we focus on the binary case $\mathcal{T} = \{-1, 1\}$.

The structure of the network of classifiers specifies which classifiers can directly influence each other. The interactions are specified by a graph $G = (V, E)$ where the nodes $V$ are fragments to be classified and the edges $E$ indicate possible dependencies. Formally, a CRF specifies Markov independence properties between the inputs and outputs as follows [4]:

**Definition 2.1** *Random variables* $(\mathbf{x}, \mathbf{y})$ *are a* conditional random field (CRF) *if, when conditioned on* $\mathbf{x}$*, all* $y_i$ *obey the Markov property with respect to the graph: namely* $P(y_i|\mathbf{x}, \mathbf{y}_{V-i}) = P(y_i|\mathbf{x}, \mathbf{y}_{\mathcal{N}_i})$*, where* $\mathbf{y}_{V-i}$ *is the set of all nodes in* $G$ *except the node* $i$*, and* $\mathcal{N}_i$ *is the set of neighbors of the node* $i$ *linked with edges in* $E$*.*

Unlike traditional Markov random fields (MRFs) which are generative models, CRFs only model the conditional distribution $P(\mathbf{y}|\mathbf{x})$ and do not explicitly model the marginal $P(\mathbf{x})$. Note that the individual labels $y_i$ are globally conditioned on the whole observation $\mathbf{x}$ in CRFs. This global conditioning allows very flexible features that can capture long-distance dependencies, arbitrary correlation, and practically any aspect of a drawing. In contrast, traditional generative models require features to be conditionally independent given the labels.

### 2.2. Individual Classification

At each node in the graph there is a classification to be made. In a CRF with no interactions ($E = \varnothing$), we can apply a classifier independently to each node $i$ and assign (a two-class) label probability

$$P_i(y_i|\mathbf{x}, \mathbf{w}) = \frac{1}{Z(\mathbf{w})} \Psi(y_i \mathbf{w}^{\mathrm{T}} \mathbf{g}_i(\mathbf{x})). \quad (1)$$

Here, $\mathbf{g}_i(\mathbf{x})$ are features associated at site $i$, usually based on observations in a local neighborhood, but potentially also dependent on global properties of $\mathbf{x}$. The features are linearly weighted by $\mathbf{w}$, and then fed through a nonlinearity $\Psi$ and normalized to sum to 1 by $Z(\mathbf{w})$. Common choices for the nonlinearity are $\Psi = \exp$, in which case we obtain a logistic classifier. Alternatively, we will use the probit function, which is the cumulative distribution of a Gaussian: $\Psi(a) = \int_{-\infty}^{a} N(x; 0, 1)\, \mathrm{d}x$.

### 2.3. Joint Modeling

When we want to model node interactions indicated by edges $E$ in $G$, we must look at the joint distribution $P(\mathbf{y}|\mathbf{x})$.

The Hammersley-Clifford theorem shows that the CRF conditional distribution $P(\mathbf{y}|\mathbf{x})$ can be written as a normalized product of potential functions on cliques of the graph (i.e., complete subgraphs of the graph). We will employ two types of potentials. Firstly, for each node we introduce a site potential $\Phi_i(y_i, \mathbf{x}; \mathbf{w})$, which measures the compatibility of one label with its associated ink, as a function of model parameters $\mathbf{w}$. Secondly, for each edge there is an interaction potential $\Omega_{i,j}(y_i, y_j, \mathbf{x}; \mathbf{v})$, which measures the compatibility between two neighboring labels, depending on their associated ink and parameters $\mathbf{v}$. We collect all parameters in $\boldsymbol{\theta} = [\mathbf{w}\ \mathbf{v}]$. Now, the CRF defines the joint label probability as

$$P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \prod_{i \in V} \Phi_i(y_i, \mathbf{x}; \boldsymbol{\theta}) \prod_{(i,j) \in E} \Omega_{i,j}(y_i, y_j, \mathbf{x}; \boldsymbol{\theta}) \quad (2)$$

$$\text{and } Z(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \Big( \prod_{i \in V} \Phi_i(y_i, \mathbf{x}; \boldsymbol{\theta}) \prod_{(i,j) \in E} \Omega_{i,j}(y_i, y_j, \mathbf{x}; \boldsymbol{\theta}) \Big)$$

$Z(\boldsymbol{\theta})$ is a normalizing constant known as the partition function.

Both types of potentials use a linearly weighted combination of ink features passed through a nonlinearity $\Psi$:

$$\text{Site} \qquad \Phi_i(y_i, \mathbf{x}; \boldsymbol{\theta}) = \Psi(y_i \mathbf{w}^{\mathrm{T}} \mathbf{g}_i(\mathbf{x})) \quad (3)$$

$$\text{Interaction} \quad \Omega_{i,j}(y_i, y_j, \mathbf{x}; \boldsymbol{\theta}) = \Psi(y_i y_j \mathbf{v}^{\mathrm{T}} \mathbf{f}_{ij}(\mathbf{x})), \quad (4)$$

Note that the strength of interaction potentials may depend on the observation $\mathbf{x}$ through the feature $\mathbf{f}_{ij}(\mathbf{x})$. In traditional random fields the interaction potentials do not depend on observations $\mathbf{x}$. One can view the interaction potential (4) as a classifier of pairs of neighboring labels.

Importantly, we make no restrictions on the relations between features $\mathbf{g}_i(\mathbf{x})$ and $\mathbf{g}_j(\mathbf{x})$, nor on $\mathbf{f}_{ij}(\mathbf{x})$ for different sites $i$ and $j$. For example, features can overlap, be strongly correlated, and extend over long distances.

We consider two nonlinearities. Firstly, the exponential function $\Psi = \exp$, which is convenient for maximizing log-likelihood and MAP, because after simplification only the linear argument remains. Secondly, we use $\Psi$ = probit function. In both cases we can also include a label noise probability $\epsilon$, which increases robustness by considering label errors in two-class problems. Specifically,

$$\Phi_i(y_i, \mathbf{x}; \boldsymbol{\theta}) = (1 - \epsilon)\Psi(y_i \mathbf{w}^{\mathrm{T}} \mathbf{g}_i(\mathbf{x})) + \epsilon \Psi(-y_i \mathbf{w}^{\mathrm{T}} \mathbf{g}_i(\mathbf{x})), \quad (5)$$

and similarly for the interaction potential.

By combining site and interaction potentials, the CRF is effectively a network of coupled classifiers. Each site potential acts like a classifier predicting the label at one site in the graph. These predictions are then coupled by classifiers based on interaction potentials.

We have illustrated CRFs for two classes, however, the multiple class case can be reduced to two classes by absorbing the labels into the feature functions $\mathbf{g}_i(\mathbf{x}, y_i)$ and $\mathbf{f}_{ij}(\mathbf{x}, y_i, y_j)$ with little change.

## 3. Training CRFs

We train the CRFs in a discriminative way. Given a set of training data, we find the parameters $\boldsymbol{\theta} = [\mathbf{w}\ \mathbf{v}]$ that maximize conditional MAP

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}} \log P(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) = \text{argmax}_{\boldsymbol{\theta}} \log P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) \quad (6)$$

We assign independent Gaussian priors to the parameters, $P(\boldsymbol{\theta}) = N(\boldsymbol{\theta}; 0, \sigma^2 \mathbf{I})$. For exponential nonlinearities and $\epsilon$=0, the $\log P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) P(\boldsymbol{\theta})$ simplifies to $\mathcal{L} =$

$$\sum_{i \in V} y_i \mathbf{w}^{\mathrm{T}} \mathbf{g}_i(\mathbf{x}) + \sum_{(i,j) \in E} y_i y_j \mathbf{v}^{\mathrm{T}} \mathbf{f}_{ij}(\mathbf{x}) - \log Z(\boldsymbol{\theta}) - \frac{\|\boldsymbol{\theta}\|^2}{2\sigma^2} \quad (7)$$

(plus a constant), and its gradients w.r.t. $\mathbf{w}$ and $\mathbf{v}$ are respectively

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mathbf{w}} : \sum_{i \in V} y_i \mathbf{g}_i(\mathbf{x}) - \Big\langle \sum_{i \in V} y_i \mathbf{g}_i(\mathbf{x}) \Big\rangle_{P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})} - \frac{\mathbf{w}}{\sigma^2},$$

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mathbf{v}} : \sum_{(i,j) \in E} y_i y_j \mathbf{f}_{ij}(\mathbf{x}) - \Big\langle \sum_{(i,j) \in E} y_i y_j \mathbf{f}_{ij}(\mathbf{x}) \Big\rangle_{P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})} - \frac{\mathbf{v}}{\sigma^2},$$

The angle brackets denote expectations with respect to the current model distribution. Since only sums of single $y_i$ or pairs $y_i y_j$ occur in the expectations, only individual marginals $P(y_i|\mathbf{x}, \boldsymbol{\theta})$ and pairwise marginals $P(y_i, y_j|\mathbf{x}, \boldsymbol{\theta})$ are required.

For probit $\Psi$ nonlinearities with the label noise model, the gradient has a similar form

$$\frac{\mathrm{d}\mathcal{L}_\psi}{\mathrm{d}\mathbf{w}} : \sum_{i \in V} q_i y_i \mathbf{g}_i(\mathbf{x}) - \Big\langle \sum_{i \in V} q_i y_i \mathbf{g}_i(\mathbf{x}) \Big\rangle_{P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})} - \frac{\mathbf{w}}{\sigma^2},$$

$$\text{where } q_i = \frac{N(y_i \mathbf{w}^{\mathrm{T}} \mathbf{g}_i(\mathbf{x}))}{\Psi(y_i \mathbf{w}^{\mathrm{T}} \mathbf{g}_i(\mathbf{x})) + \frac{\epsilon}{1-2\epsilon}}$$

and likewise for the gradient for the interaction parameters $\mathbf{v}$.

For both exponential and probit nonlinearities, the log-likelihood is concave when the label noise $\epsilon = 0$. Thus, gradient ascent is guaranteed to find a global maximum. The quasi-Newton technique BFGS [5] converges in 50-100 iterations in our application.

The computational cost is dominated by calculating the partition function $Z(\boldsymbol{\theta})$ and the marginals $P(y_i|\mathbf{x}, \boldsymbol{\theta})$ and $P(y_i, y_j|\mathbf{x}, \boldsymbol{\theta})$. In general, an exact calculation is exponential in the number of nodes in the graph, but fortunately our graphs are sparsely connected. In this case, the junction tree algorithm is feasible on the triangulated graph [2].

Our ink graphs have a tree width typically less than 5, and require around 5000 FLOPS to calculate a complete set of marginals and the partition function. For more densely connected graphs, approximate inference such as loopy belief propagation may be necessary.

## 4. Inference on CRFs

Unlike traditional classification problems, where we find the probability of a single label given an input, a CRF assigns a joint probability to a configuration of labels given an input $\mathbf{x}$ and parameters $\boldsymbol{\theta}$. We are typically interested to find either the maximum a posteriori (MAP) or maximum marginal (MM) solution:

$$\mathbf{y}^{\text{MAP}} = \text{argmax}_{\mathbf{y}} P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) \quad (8)$$

$$y_i^{\text{MM}} = \text{argmax}_{y_i} P(y_i|\mathbf{x}, \boldsymbol{\theta}), \quad \forall i \in V. \quad (9)$$

The MAP solution finds a globally compatible label assignment, whereas the max marginal solution will greedily choose the most likely individual labels, which may disagree with each other (even though they arise from a joint distribution). However, in a practical recognition scenario we like to minimize the number of individually mislabeled segments, hence the MM criterion is appropriate and usually performs slightly better than MAP.

To find the MM solution we require individual marginals, which we calculate exactly, as done during training. The MAP solution can also be calculated exactly using the max-product algorithm applied to the junction tree. Again, approximate techniques may be necessary for dense graphs with loops [7].

## 5. Application to Ink Classification

Here we apply CRFs to online ink classification, specifically to discriminating between containers and connectors in drawings of organization charts. Context is exploited by joint classification where the labeling of one ink fragment influences the labels of the others.

We break the task into three steps:

1. Subdivision of pen strokes into fragments,
2. Construction of a CRF on the fragments,
3. Training and inference on the random field.

The input is electronic ink recorded as sampled locations of the pen, and collected into *strokes* separated by pen-down and pen-up events. In the first step, strokes are divided into simpler components called *fragments*. Fragments should be small enough to belong to a single container or connector. In contrast, strokes can occasionally span more than one shape, for example when a user draws a container and a connector without lifting the pen. We choose fragments to be groups of ink dots within a stroke that form straight line segments (within some tolerance) (Figure 1).

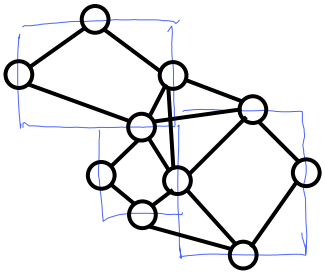

**Figure 2. The conditional random field superimposed on part of the chart from Figure 1. There is one node (circled) per fragment, and edges indicate interaction potentials between neighboring fragments.**

In the second step, we construct a conditional random field on the fragments. Each ink fragment is represented by a node in the graph (Figure 2). The node has an associated label variable $y_i$, which takes on the values -1 (container) or 1 (connector). CRF potential functions $\Phi$ and $\Omega$ quantify how compatible labels are with the underlying ink and with neighboring labels. Weights in the potential functions characterize the exact dependence of labels with the ink.

The site potentials refer to the label $y_i$ of a single fragment and its ink context. The context can be any subset of all ink $\mathbf{x}$, but typically only neighboring fragments are included. Note that site potentials are already very powerful compared to approaches that model each fragment of ink independently.

Interaction potentials model whether a pair of fragments prefer the same or differing labels. These potentials can depend on features of the pair, such as the nearest distance or angle between the fragments. Again, they can also take other fragments into account as context, typically neighboring fragments.

Our approach is to compute many redundant low-level ink features, and represent them in potentials in the random field. The CRF algorithm then learns which features or combinations of features that are discriminative for the task.

Our two simplest features are the length and orientation angle of an ink fragment. These are encoded in site potentials. Secondly, we consider the context of a single fragment. We calculate the histogram of distances and relative angles to neighboring fragments, and use these as vector-valued features. Next, for interaction potentials, we compute features depending on pairs of fragments $i$ and $j$. These include the distance and angle between the fragments, and temporal features such as whether the pen was lifted in between them.

Finally, we include template features that detect simple perceptual relations. We employ domain-knowledge to capture parts of organization charts. We employ a basic corner and a T-junction feature, a container-side feature that checks whether corners are present on both ends of a fragment, and an alignment measure of whether two fragments are parallel and aligned. Some of these features yield real number values, but most are binary. Lastly, we include a bias feature that is always 1. For other recognition tasks, appropriate features can be added easily.

## 6. Experiments and Discussion

We asked 17 subjects to draw given organization charts on a TabletPC device capturing online handwriting. The given charts consisted of rectangular containers and connectors made from line segments. We focused on graphical elements, and removed any text.

The pen strokes were subdivided yielding a database of 1000 fragments, which we split into training sets drawn by half of the subjects, and test sets drawn by the other half. We built a CRF with site potential functions for each fragment, and interaction potential between all pairs of fragments that were within 5mm of each other, resulting in 3000 interaction potentials.

Since the undirected graphical models generated from the organization charts were sparse, triangulation yielded junction trees with low tree width and we trained with BFGS. The template features (T-junction) were weighted most heavily by the classifier. For inference, we ran the max-product algorithm to determine global MAP and max-margin solutions. Priors for the weight parameters were set to $\sigma$=2 and a no label error model $\epsilon$=0 was used. However, $\epsilon > 0$ actually gives significantly better accuracy (detailed experiments are in progress).

We measured the performance on different types of organization charts: type A and B (the latter exemplified in Figure 4), as well as a mixed set with four chart types. The first and third rows in Table 1 show test errors for classification using only the site potentials, i.e., individual classification of fragments. The second and fourth rows give errors for conditional random fields that propagate information from both site and interaction potentials. These results were produced from max-marginals, which gave almost identical results to MAP, but MM appears better when label error $\epsilon$ is non-zero. Both MAP and MM labeling takes less than one second per drawing.

A typical result is shown in Figure 3. To the left we see the results of a CRF employing only site potentials (individual classification.) There are two ambiguous rectangles created from fragments 18-21-22-29 and 14-16-17. The site potentials misclassify fragments 13, 14, 21 and 22. The CRF with interaction potentials resolves the ambiguity using context and correctly classifies all fragments. Similar benefits arise for other charts (Figure 4).

| | Nonlin. | Type A | Type B | Mixed |
|---|---|---|---|---|
| Individual | exp | 4.4% | 12.1 % | 10.2% |
| Joint | exp | 0% | 3.5 % | 7.7% |
| Individual | probit | 4.4% | 11.6 % | 9.3% |
| Joint | probit | 0% | 3.5 % | 6.2% |

**Table 1. Classification errors for individual and joint classification for two nonlinearities.**

## 7. Conclusion

We have demonstrated that exploiting context can significantly improve recognition accuracy. We proposed joint classification with conditional random fields for analysis of hand-drawn diagrams. The framework is general and allows flexible features. The joint probability output enables us to find not only the most likely labeling, but also to rank the top labelings, also after further user corrections of individual parts. The technique can be applied in many other handwriting and image recognition tasks.

In the future, we will consider densely connected ink graphs using approximate inference or intelligent choice of graph structure. Also, we can kernelize CRF potential functions to enhance their modeling power. Moreover, we will apply CRFs to labeling more than two classes. A challenging problem is to apply CRFs to hierarchical parsing of a 2D scene.

## Acknowledgments

We are grateful to Thomas Minka, Michel Gangnet, Christopher Bishop, Philip Cowans, and Hannah Pepper for valuable discussions, great inference software and help with data collection and experiments.

## References

[1] J. Coughlan and S. Ferreira. Finding deformable shapes using loopy belief propagation. In *European Conf. on Computer Vision*, 2002.

[2] F. Jensen. *Bayesian Networks and Decision Graphs*. Springer, 2001.

[3] S. Kumar and M. Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *IEEE Intl. Conf. on Computer Vision*, 2003.

[4] J. Lafferty, A. McCallum, and F. Pereira. Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. In *Intl. Conf. Machine Learning*, pages 282–289, 2001.

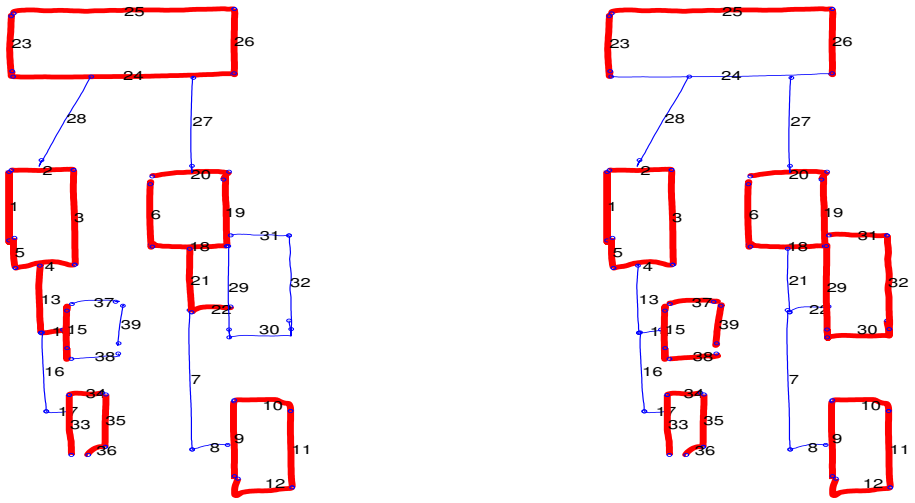

**Figure 3. Classification of a chart from the mixed group. Left: individual classification with a CRF employing site potentials only, hence no modeling of dependencies across nodes. Right: joint classification, using a full CRF with site and interaction potentials. Containers are shown in bold and connectors are thin lines.**

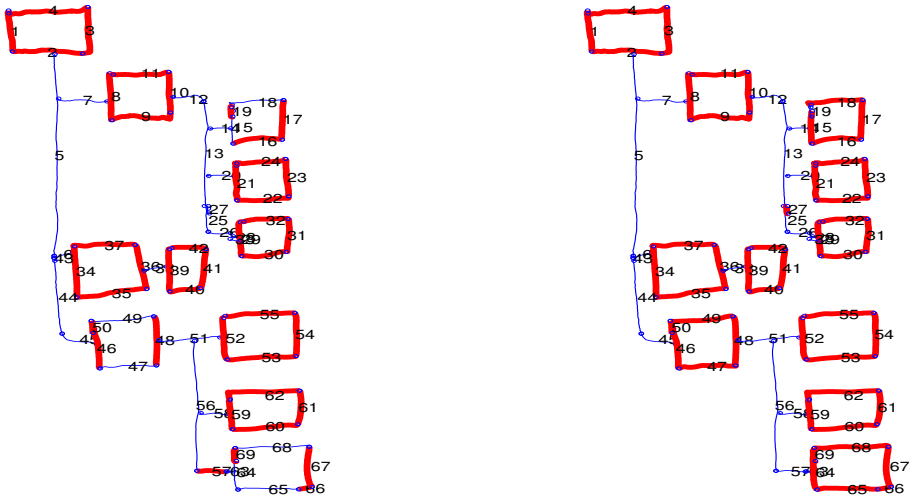

**Figure 4. Classification of a chart of Type B. Individual and joint classification (left and right, respectively). Joint classification corrects individually misclassified fragments, except fragment 27.**


[5] R. Malouf. A comparison of algorithms for maximum entropy parameter estimation. In *6th Workshop on Computational Language Learning (CoNLL)*, 2002.

[6] Z. Tu, X. Chen, A. Yuille, and S.-C. Zhu. Image parsing: Segmentation, detection, and object recognition. In *Intl. Conf. Computer Vision ICCV*, 2003.

[7] Y. Weiss. Belief propagation and revision in networks with loops. Technical Report 1616, MIT AI Lab, 1997.