A Two-Step Soft Segmentation Procedure for MALDI Imaging Mass Spectrometry Data

Ilya Chernyavsky[1], Theodore Alexandrov[2,3], Peter Maass[2], and Sergey I. Nikolenko[1,4]

1 **St. Petersburg Academic University, ul. Khlopina, d. 8, korp. 3, St. Petersburg, 194021, Russia,** ilya.chernyavsky@gmail.com
2 **Center for Industrial Mathematics, University of Bremen, Bibliothekstr. 1, 28359 Bremen, Germany,** theodore@uni-bremen.de
3 **Steinbeis Innovation Center SCiLS, Richard-Dehmel-Str. 69, 28211 Bremen, Germany**
4 **Steklov Mathematical Institute, nab. r. Fontanka, 27, St. Petersburg, 191023, Russia,** sergey@logic.pdmi.ras.ru

Abstract

We propose a new method for soft spatial segmentation of matrix assisted laser desorption/ionization imaging mass spectrometry (MALDI-IMS) data which is based on probabilistic clustering with subsequent smoothing. Clustering of spectra is done with the Latent Dirichlet Allocation (LDA) model. Then, clustering results are smoothed with a Markov random field (MRF) resulting in a soft probabilistic segmentation map. We show several extensions of the basic MRF model specifically tuned for MALDI-IMS data segmentation. We describe a highly parallel implementation of the smoothing algorithm based on GraphLab framework and show experimental results.

1998 ACM Subject Classification I.4.6 Segmentation, J.3 Life and Medical Sciences, J.2 Physical Sciences and Engineering, H.2.8 Database Applications

Keywords and phrases MALDI imaging mass spectrometry, hyperspectral image segmentation, probabilistic graphical models, latent Dirichlet allocation, Markov random field

Digital Object Identifier 10.4230/OASIcs.GCB.2012.39

1 Introduction

1.1 MALDI-imaging mass spectrometry and the segmentation problem

Untargeted spatially-resolved biochemical analysis of biological tissue sections using imaging mass spectrometry (IMS) is an emerging field of biochemistry which has received considerable attention over the last decade [11,16,35], with the most popular IMS technique called matrix-assisted laser desorption/ionization (MALDI-IMS) or MALDI-imaging [9,32]. IMS acquires a set of mass spectra in a rasterized way across the surface of a prepared tissue section; one mass spectrum corresponds to one pixel. A mass spectrum is a vector of intensities associated with mass-to-charge (m/z) values representing in a semi-quantitative way the abundances of ions separated by their m/z-values; a high intensity of a mass spectrum at an m/z-value represents a high abundance of an ion of this m/z. In MALDI mass spectrometry, an m/z-value is usually interpreted as the molecular mass, since ions with charge $+1$ prevail. An IMS dataset is large, normally with around 100×100 spectra acquired across the section with each spectrum having around 10000 intensity values. An IMS dataset can be considered as a hyperspectral image whose image channels correspond to m/z-values.

German Conference on Bioinformatics 2012 (GCB'12).
Editors: S. Böcker, F. Hufsky, K. Scheubert, J. Schleicher, S. Schuster; pp. 39–48

OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Automatic mining of such massive hyperspectral imaging data is essential for applications of the IMS technique to biomedical challenges. One of the established methods of MALDI-IMS data mining is spatial segmentation that allows one to represent a hyperspectral MALDI-IMS dataset with just one image, the so-called segmentation map. In this map, pixels of similar mass spectra are pseudo-colored with the same color. There are several methods proposed for spatial segmentation of MALDI-IMS data based on clustering of mass spectra. Spatial segmentation of MALDI-IMS data is challenging because of the size of the datasets and strong pixel-to-pixel variation inherent to the sample preparation and acquisition techniques used for MALDI-IMS. It has been shown that the best-performing segmentation methods reduce pixel-to-pixel variation, e.g. by means of incorporating spatial relations between pixels into the distance function [2].

1.2 Results and structure of the paper

We propose a two-step procedure for spatial segmentation of MALDI-IMS data. On the first step, we cluster MALDI-IMS spectra with the probabilistic graphical model called Latent Dirichlet Allocation (LDA) [8]. LDA has recently obtained recognition in text mining because it can remain scalable and efficient while extracting hidden features ("topics") from a set of documents. In the context of MALDI-IMS data segmentation, LDA has the following advantages as compared to other published methods: (i) reduced memory requirements as compared to hierarchical clustering recommended by [12]; (ii) reduced complexity as compared to high-dimensional discriminant clustering recommended by [1]; (iii) most importantly, additional information provided by LDA as compared to other clustering methods. This information associates each data point (pixel) in the LDA model with the probabilities to belong to each cluster and, moreover, the LDA model learns a distribution of the "words" (m/z values) to be generated by different topics; this "soft" clustering is better suited for further analysis such as extracting characteristic masses for each cluster. We describe our clustering procedures in Section 2.

On the second step, we smooth soft clustering results provided by LDA with a probabilistic smoothing model which we base upon Markov random fields. LDA results are convenient to use in such a probabilistic setting. We introduce several extensions to the model that let us enhance MRF smoothing results specifically for MALDI-imaging data processing; in particular, we learn the weights of the MRF network with an EM-like procedure. Moreover, we have developed a fast highly parallel implementation of MRF smoothing for clustering results based on the GraphLab parallelization paradigm. These results are described in Section 3. In Section 4, we describe the dataset and show experimental results of our segmentation procedures. Section 5 concludes the paper.

2 Clustering of MALDI-IMS spectra

2.1 LDA

Latent Dirichlet allocation (LDA) is a clustering model that was originally intended for use with natural language text processing [8, 24]. LDA takes the naive Bayes model one step further: while naive Bayes has only one latent variable, namely the topic, and words are conditionally independent given a topic, LDA view documents as *mixtures* of several topics. In essence, LDA is a hierarchical Bayesian model: (i) on the first level, it has a mixture whose components correspond to the "topics"; (ii) on the second level, a multinomial variable with Dirichlet prior that corresponds to the "distribution of topics" in a document.

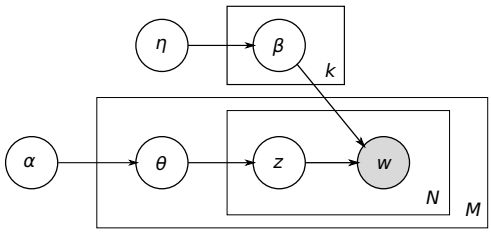


Figure 1 LDA graphical model.

This gives rise to the following generative model (see Fig. 1): (1) choose the number of words $N \sim p(N \mid \xi)$; (2) choose the topic distribution $\theta \sim \mathrm{Dir}(\alpha)$; (3) for each word w_n, $n = 1..N$: (i) choose a topic $z_n \sim \mathrm{Mult}(\theta)$; (ii) choose a word $w_n \sim p(w_n \mid z_n, \beta)$. The joint distribution in the LDA model (for a fixed number of topics k and parameters $\beta_{ij} = p(w^j = 1 \mid z^i = 1)$) is $p(\theta, \mathbf{z}, \mathbf{w}, N \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) = p(N \mid \xi)p(\theta \mid \boldsymbol{\alpha}) \prod_{n=1}^{N} p(z_n \mid \theta)p(w_n \mid z_n, \boldsymbol{\beta})$. Inference in LDA can be done, for example, with variational methods or with Gibbs sampling.

The LDA model has had success in text processing, but its applications are not limited to information retrieval problems. In this work, we have applied LDA to the clustering of the spectra. Note that in MALDI-imaging, one spectrum is acquired at one pixel. We propose to use individual pixels of the image as "documents" and masses corresponding to peaks in the spectrum as "words". The results of applying LDA to such "documents" are topic distributions in each pixel of the image and a mass distribution for every topic. The number of topics is a parameter defined in advance. Fig. 3a shows the results of a direct application of LDA with 10 topics to the original dataset. Each color corresponds to a topic, and a pixel represents a topic with maximal posterior probability. Fig. 3b shows a soft segmentation map representing the second, fourth, and tenth clusters in the segmentation map from Fig 3a. A soft segmentation map is an RGB-image where three clusters are encoded with red, blue, and green channel respectively. Intensities of each channel represent probabilities of a pixel to belong to the corresponding cluster. For example, a highly red pixel has high probability to be in the second cluster; a blue pixel probably belongs to the fourth cluster, and a green pixel probably belongs to the tenth cluster. Compared to a standard segmentation map (Fig. 3a), a soft segmentation map, although restricted to visualize only three clusters, provides more detailed information on probability distribution of pixels among the clusters. In particular, it shows pixels with comparable probabilities of belonging to different clusters.

2.2 k-means

For comparison, we have implemented k-means as a simple yet effective clustering algorithm [21, 23]. In k-means, we fix the number of clusters and initialize their centers (usually at random dataset points that are chosen to be relatively far from each other). Then k-means proceeds in alternating steps: (1) assign points to clusters; each point is assigned to the nearest cluster center; (2) move cluster centers to the centers of mass for the points assigned to them.

3 Smoothing probabilistic clustering results

3.1 Markov random fields

An *undirected graphical model* (or *Markov random field*) [5, 22, 27, 29] is a representation of a complex probability distribution factorized into a product of *potentials*; the model consists of an undirected graph with vertices corresponding to random variables $X_1, \ldots, X_k$ and a potential for each maximal clique in this graph, i.e., a nonnegative function of the variables; the joint probability distribution corresponding to the model is the product of potentials: $p(X_1 = x_1, \ldots, X_k = x_k) = \frac{1}{Z} \prod_C \psi_C(x_C)$, where C are the maximal cliques, ψ_C

are nonnegative functions (potentials), and Z is the normalization constant often called the *partition function*. Since ψ_C are nonnegative, they are often represented in the exponential form: $\psi_C(x_C) = \exp(-E_C(x_C))$; E_C are often called *energy functions*, and the exponential representation of the total energy $p(X) = \exp(-\sum_C E_C(x_C))$ is called the *Boltzmann distribution* (this field borrows much of its terminology from statistical physics).

This definition implies a simple conditional independence statement: $p(x_i,\ x_j \mid x_{k\neq i,j}) = p(x_i \mid x_{k\neq i,j})p(x_j \mid x_{k\neq i,j})$ if x_i and x_j are not connected by an edge. Moreover, $p(\mathcal{X},\ \mathcal{Y} \mid \mathcal{Z}) = p(\mathcal{X} \mid \mathcal{Z})p(\mathcal{Y} \mid \mathcal{Z})$ if every path from $\mathcal{X}$ to $\mathcal{Y}$ goes through $\mathcal{Z}$, where $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are disjoint subsets of random variables.

Undirected graphical models have a rich history; in particular, undirected graphical models have been widely used for image denoising and similar problems, such as image segmentation, which is the primary motivation for our current study [13, 25, 28, 29]. Image denoising was one of the first applications for undirected graphical models [3, 4, 14].

In image processing problems, undirected models usually appear in the form of rectangular grids of pixels, with potentials on the edges intended to smooth the transitions; i.e., potentials usually serve to bring the neigboring nodes closer together. A simple yet useful model on a rectangular grid is the *Ising model*: energy functions on the edges of the grid are given by $\epsilon_{i,j}(x_i, x_j) = -w_{i,j}x_ix_j$, and the resulting joint distribution is $p(x_1, \ldots, x_n) = \frac{1}{Z}\exp\left(\sum_{(i,j)\in E} w_{i,j}x_ix_j + \sum_i u_ix_i\right)$. Inference in undirected graphical models can be done via *belief propagation*. In the basic message passing algorithm, a node collects messages from other nodes and sends out the marginalized results; for a detailed description of belief propagation, we refer to [5, 27]. When the underlying graph of the model contains cycles (and a grid model certainly does), belief propagation is not guaranteed to converge but when it converges, it does find a local minimum of the total energy function; this version of the algorithm is called *loopy belief propagation*.

These models also allow for a relatively simple approximate inference algorithm based on Gibbs sampling [5, 10, 14, 30]. For an Ising model with weights $w_{i,j}$ and u_i, the Gibbs sampler sequentially updates all variables by fixing all values of the variables except one, say x_i, and then sampling x_i from the conditional distribution: $x_i \sim p(x_i \mid \mathbf{x}_{-i})$. The conditional distribution is easy to compute in this case:

$$p(x_i = k \mid \mathbf{x}_{-i}) = \frac{p(x_i = k, \mathbf{x}_{-i})}{\sum_s p(x_i = s, \mathbf{x}_{-i})} = \frac{\prod_{j\in \mathrm{nei}(i)} \psi_{ij}(x_i = k, x_j)}{\sum_s \prod_{j\in \mathrm{nei}(i)} \psi_{ij}(x_i = s, x_j)},$$

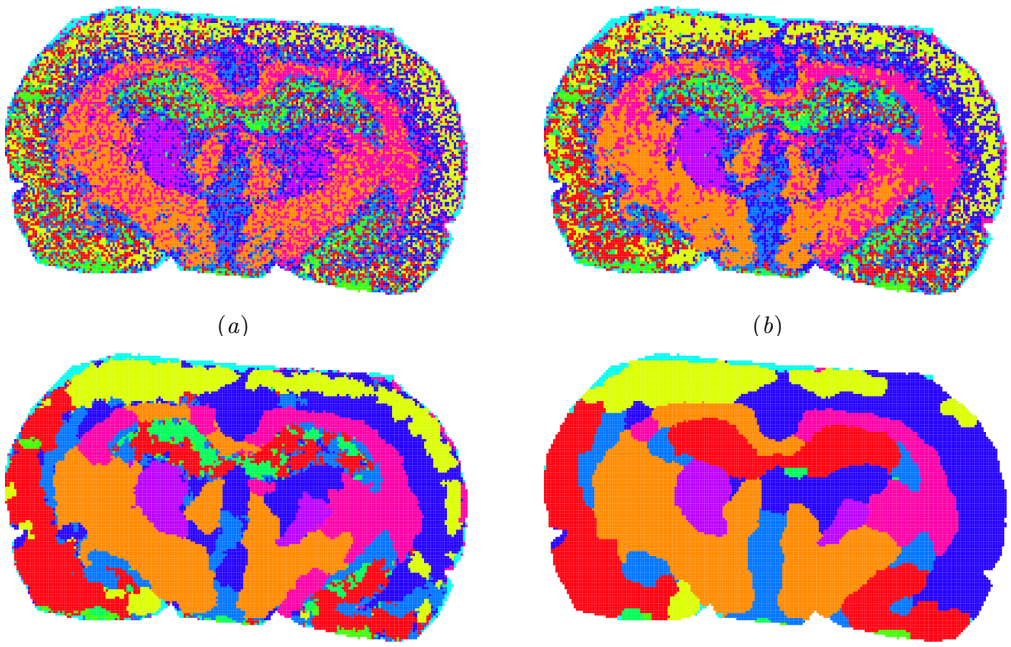
where $\mathrm{nei}(i)$ is the neighbors set of the i^{th} vertex. Gibbs sampling is a special case of the Metropolis-Hastings algorithm for Monte-Carlo Markov chain sampling.

3.2 Our two-step approach

Markov random fields have already been successfully applied in the field of imaging mass spectrometry [15]. In this paper, we present a two-stage procedure for the segmentation of MALDI-imaging data. The first step is based on latent Dirichlet allocation, and the second step is based on Markov random fields used for denoising.

In the first step we learn the LDA model from the data. To do this, we need to define "documents" and their contents ("words"). The method we propose treats each pixel of a MALDI-imaging dataset as a document and considers the masses with high intensity values (peaks) in the associated spectrum as the words of the document. The result of this step consists of topic distributions θ_i for each pixel and word (peak) distributions for each topic.

In the second step, we use a Markov random field similar to the Ising model in order to denoise (smooth) the segmentation map. The model has two types of factors: binary factors



(a) (b) (c) (d)

Figure 2 Segmentation maps obtained with k-means and MRF: (a) no denoising; (b) $w = 0.3$; (c) $w = 0.6$; (d) $w = 0.9$.

that correspond to the edges and unary factors that correspond to the vertices. A binary factor $\phi_{i,j}(t_1, t_2)$ takes the value $\exp(w)$ whenever $t_1 = t_2$ and $\exp(-w)$ otherwise. These factors are responsible for the smoothing part of the model. The unary factors are responsible for the incorporation of topic distributions found on the previous step; we initialize unary factors simply as $u_i(t) = \theta_i(t)$, where θ_i is the topic distribution for pixel i and $\theta_i(t)$ is the probability of topic t.

Previously, we developed other two-step approaches, where denoising was applied prior to clustering [1] and where denoising was embedded into clustering [2]. In this paper, we considered the potential of the two-step approach where the denoising is applied after probabilistic clustering. Note that our approach is different from calculating the segmentation map first and then denoising it for better visualization that might lead to losing details during denoising. In line with [15], we first reduce the dataset to k images, where each image represents the probabilities of pixels to belong to the corresponding topic. Then, the Markov random field exploits all k probabilities calculated for one pixel to decide on the assignment of this pixel to a specific cluster of the segmentation map.

For comparison, we have also implemented an approach in which the first step is based on the k-means algorithm. In this approach, we consider a MALDI-imaging dataset as a set of vectors and apply k-means clustering to it. The k-means algorithm outputs a set of cluster centers. Next we use a Markov random field for segmentation map smoothing. Binary factors are the same as above while unary factors look like $u_i(t) = 1/||s_i - c_t||$, where s_i is the spectrum associated with pixel i and c_t is the center of the t^{th} cluster.

Thus, in both cases the models try to smooth the image and at the same time preserve the segmentation map found on the first step. The use of LDA model on the first step not only increases the quality of the resulting segmentation but also provides additional information

concerning the structure of segments through the distributions on topics for each pixel and distribution on words, i.e., masses, for each topic.

3.3 MRF extensions for MALDI-imaging

One problem with MRF smoothing for MALDI-IMS data is that the prior distribution induced by the smoothing factors with constant weights might not be flexible enough to represent the complex anatomical structures of the brain. In this section, we present two modifications of the MRF approach that significantly improve segmentation results for MALDI-IMS data.

The first modification is based on the use of Kullback-Leibler divergence. Instead of a constant weight w, we set the smoothing weight between neighboring pixels i and j depending on the Kullback-Leibler divergence between the topic distributions θ_i and θ_j that correspond to these pixels. In particular, we decrease the value of the smoothing weight when the Kullback-Leibler divergence exceeds the chosen threshold. This modification allows for different smoothing weights depending on the similarity of the pixels which results in reducing of the noise while preserving the details of the picture.

The second modification works through a simple version of the EM algorithm. Here, we use a more complicated form of the smoothing binary factor. We assume that the model has a set of parameters $\{\lambda_{t_1,t_2}\}$ and that the binary factor $\phi_{i,j}(t_1, t_2)$ has the following form: $\phi_{i,j}(t_1, t_2) = \exp(w\lambda_{t_1,t_2})$. Then we use one iteration of the EM algorithm to estimate the parameters. To do this, we first initialize binary factors with unit factors, $\lambda_{t_1,t_2} = 1$ when $t_1 = t_2$ and $\lambda_{t_1,t_2} = -1$ otherwise. After that, we use the Markov random field to get the denoised segmentation map. Then we reestimate λ_{t_1,t_2} as $\lambda_{t_1,t_2} = N_{t_1,t_2}/N$, where N_{t_1,t_2} is the number of neighboring pixel pairs with values t_1 and t_2 and N is the number of edges. After that, we normalize each row in the matrix of λ_{t_1,t_2} values and use the Markov random field with updated binary factors to get the final segmentation map.

3.4 Parallel implementation

MRF inference algorithms take up the greater portion of the running time in our two-step system. We have developed a highly parallel implementation of loopy belief propagation based on GraphLab, a recently developed framework for parallelizing large-scale machine learning tasks [26]. In the GraphLab paradigm, the data is represented as a graph, and the algorithm is represented as two basic routines: (i) Update runs in a single node and can use the data in this node, on its adjacent edges, and on its neighboring vertices; (ii) Sync collects the results from several subgraphs and propagates them further up. GraphLab can be viewed as a generalization of MapReduce. However, while MapReduce requires data subsets for the Map function to be completely independent, GraphLab allows them to form a complex highly connected graph, requiring only that the graph is relatively sparse (otherwise, there will be no speed improvement from parallelization).

GraphLab guarantees that during an Update procedure the data that it uses cannot be changed by any other Update, thus ensuring that there are no deadlocks and data races. There are several levels of data independence: an Update in a node may lock either its entire neighborhood (adjacent nodes and incident edges), incident edges only, or no other graph elements. In the case of loopy belief propagation in MRF, it suffices to lock incident edges since the message passing algorithm changes messages on the edges but does not change anything in other nodes. The resulting system shows promising speed and accuracy; it performs loopy belief propagation on an MRF until convergence in time which rapidly decreases with the number of cores; see Fig. 4 for a comparison.

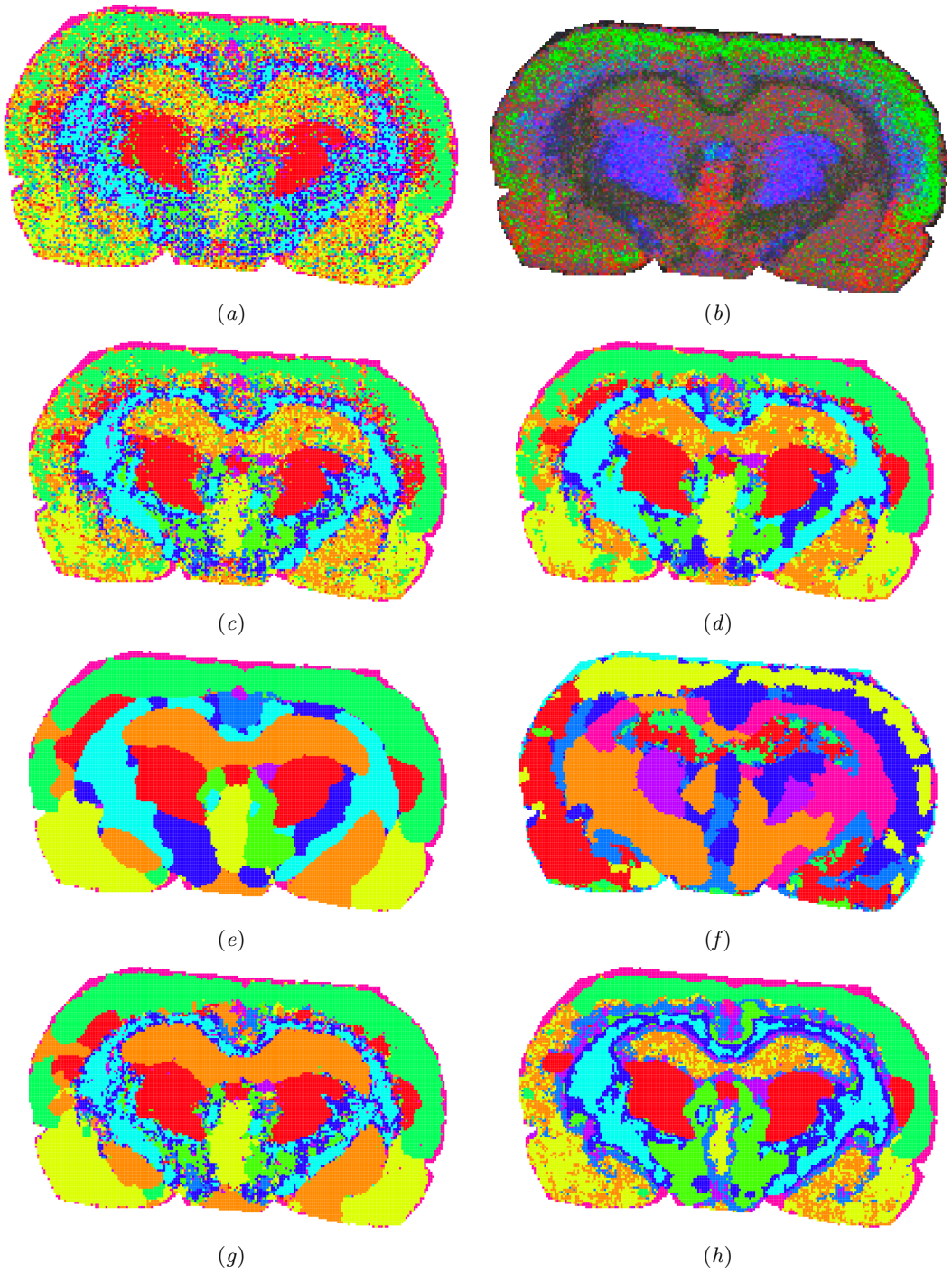


Figure 3 Segmentation maps obtained with LDA and MRF. (*a*) LDA clustering results; *(b)* a soft segmentation map for three clusters; (*c*) MRF smoothing with $w = 0.4$; (*d*) $w = 0.55$; (*e*) $w = 0.7$; (*f*) Segmentation map obtained with k-means ($w = 0.6$); (*g*) Kullback-Leibler modification; (*h*) EM algorithm modification.

4 Experimental results

Samples preparation and MALDI-IMS measurements are described in detail in [1]. Shortly, the cryosections of $10\mu m$ thickness were cut on a cryostat, transferred to a conductive indium-tin-oxide coated glass slide (Bruker Daltonik GmbH, Bremen, Germany) and measured using a MALDI-TOF instrument (Autoflex III; Bruker Daltonik GmbH) using flexControl 3.0 and flexImaging 2.1 software (Bruker Daltonik GmbH). The lateral resolution was set to $80\mu m$. Preprocessing was done in ClinProTools 2.2 software (Bruker Daltonik GmbH). The spectra were baseline-corrected with the TopHat algorithm (minimal baseline width set to 10 %, the default value in ClinProTools). No normalization or binning was done. The dataset comprises 20185 spectra acquired within the area of the slice (120×201 pixels), each of 3045 data points covering the mass range 2.5–10 kDa. For examples of intensity images for several m/z-values, see [1]. Afterwards, we applied a peak picking algorithm selecting 110 m/z-values; see [1] for the details and a visualization of the peak picking results.

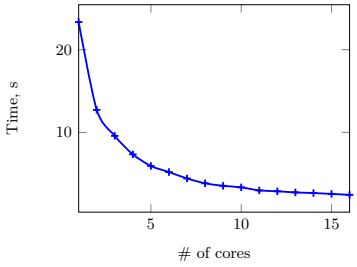


Figure 4 GraphLab implementation of loopy belief propagation on multiple cores.

Segmentation maps obtained with k-means and MRF are shown on Fig. 2; the ones obtained with LDA and MRF with varying values of w, on Fig. 3c–e. We consider the best-looking denoised picture to be the one with $w = 0.55$ (Fig. 3d). For comparison with LDA pictures, an example of segmentation map obtained with k-means is shown on Fig. 3f. We also experimented with two modifications described in Section 3.3. Results from the modification based on Kullback-Leibler divergence are shown on Fig. 3g; from the EM-based modification, on Fig. 3h. It is evident that our modification based on Kullback-Leibler divergence results in a cleaner image with less noise and at the same time better preserved details in the central part of the image. The EM-based modification produces an image that more closely resembles the anatomical structure of the brain.

5 Conclusion

In this work, we have presented a method for processing MALDI-imaging data; our method consists of two steps: probabilistic clustering with the LDA model and MRF smoothing. We have presented two novel modifications for the MRF smoothing method that significantly improve segmentation results for MALDI-imaging datasets. Further work in this direction is twofold. First, we plan to further improve clustering and denoising: improve the LDA model by learning hierarchical structures and/or correlated topics [6, 7], extend the undirected smoothing model by hidden factors, thus passing from an MRF to a restricted Boltzmann machine [17–20], incorporate ideas from segmentation algorithms based on hierarchical Dirichlet processes and hierarchical Pitman-Yor processes [31, 33, 34]. But the main direction of further work is to make the next step in analyzing the "spectrometry data cube" and apply the resulting models for a deeper analysis of mass-spectrometry data: find m/z values that are most characteristic for each segment, analyze spectra of the segments, construct most characteristic mass sections, find least characteristic sections (outliers) and so on. We hope that our approach will significantly assist us in this analysis.

Acknowledgements The authors thank Michael Becker (Bruker Daltonik GmbH, Bremen, Germany) for providing the rat brain MALDI-imaging dataset. This work was supported

by the Russian Fund for Basic Research grant 11-01-00760-a, Russian Presidential Grant for Young Ph.D.s MK-6628.2012.1, for Leading Scientific Schools NSh-3229.2012.1, Russian Governmental grant 11.G34.31.0018, and the Federal Target Programme "Scientific and Scientific–Pedagogical Personnel of Innovative Russia", 2009–2013.

References

1 Theodore Alexandrov, Michael Becker, Soeren-Oliver Deininger, Guenther Ernst, Liane Wehder, Markus Grasmair, Ferdinand von Eggeling, Herbert Thiele, and Peter Maass. Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering. *Journal of Proteome Research*, 9(12):6535–6546, 2010.

2 Theodore Alexandrov and Jan H. Kobarg. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics*, 27:i230–i238, 2011.

3 J. Besag. On spatio-temporal models and Markov fields. In *Transactions of the 7th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 47–75. Academia, 1974.

4 J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48:259–302, 1986.

5 Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

6 David M. Blei, Michael I. Jordan, Thomas L. Griffiths, and Joshua B. Tennenbaum. Hierarchical topic models and the nested chinese restaurant process. *Advances in Neural Information Processing Systems*, 13, 2004.

7 David M. Blei and John D. Lafferty. Correlated topic models. *Advances in Neural Information Processing Systems*, 18, 2006.

8 David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.

9 R. M. Caprioli, T. B. Farmer, and J. Gile. Molecular imaging of biological samples: localization of peptides and proteins using maldi-tof ms. *Analytical Chemistry*, 69(23):4751–4760, 1997.

10 G. Casella and E. I. George. Explaining the gibbs sampler. *The American Statistician*, 46:167–174, 1992.

11 Kamila Chughtai and Ron M. A. Heeren. Mass spectrometric imaging for biomedical tissue analysis. *Chemical Reviews*, 110(5):3237–77, 2010.

12 Soeren-Oliver Deininger, Matthias P. Ebert, Arne Fuetterer, Marc Gerhard, and Christoph Roecken. MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *Journal of Proteome Research*, 7(12):5230–5236, 2008.

13 Mario A. T. Figueiredo. Bayesian image segmentation using wavelet-based priors. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 437–443, Washington, DC, USA, 2005. IEEE Computer Society.

14 S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

15 Michael Hanselmann, Ullrich Köthe, Marc Kirchner, Bernhard Y. Renard, Erika R. Amstalden, Kristine Glunde, Ron M. A. Heeren, and Fred A. Hamprecht. Toward digital staining using imaging mass spectrometry and random forests. *Journal of Proteome Research*, 8(7):3558–3567, 2009. PMID: 19469555.

16 Ron M. A. Heeren, Donald F. Smith, Jonathan Stauber, Basak Kuekrer-Kaletas, and Luke MacAleese. Imaging mass spectrometry: hype or hope? *Journal of the American Society for Mass Spectrometry*, 20(6):1006–1014, 2009.

17 Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14:1771–1800, 2002.

18 Geoffrey E. Hinton. What kind of a graphical model is the brain? In *Proceedings of the* 19^{th} *International Joint Conference on Artificial Intelligence*, pages 1765–1775, 2005.

19 Geoffrey E. Hinton. A practical guide to training restricted boltzmann machines. `http://www.cs.toronto.edu/~hinton/absps/guideTR.pdf`, 2012.

20 Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

21 Paul Hudak, John Hughes, Simon Peyton Jones, and Philip Wadler. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. Berkeley, University of California Press, 1967.

22 Michael I. Jordan and Y. Weiss. Graphical models: probabilistic inference. In M. Arbib, editor, *Handbook of Neural Networks and Brain Theory*. MIT Press, 2002.

23 T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu. An efficient k–means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24:881–892, 2002.

24 S. Lacoste-Julien, F. Sha, and Michael I. Jordan. Disclda: Discriminative learning for dimensionality reduction and classification. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, Cambridge, MA, 2008. MIT Press.

25 S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Advances in Pattern Recognition. Springer, 2009.

26 Yucheng Low, Joseph Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin, and Joseph M. Hellerstein. Graphlab: A new parallel framework for machine learning. In *Proceedings of the* 26^{th} *Conference on Uncertainty in Artificial Intelligence*, 2010.

27 D. J. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

28 Patrick Perez. Markov random fields and images. *CWI Quarterly*, pages 413–437, 1998.

29 Simon Prince. *Computer Vision: Models, Learning, and Inference*. Cambridge University Press, 2012.

30 C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 2004.

31 Alex Shyr, Trevor Darrell, Michael I. Jordan, and Raquel Urtasun. Supervised hierarchical pitman-yor process for natural scene segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2281–2288, 2011.

32 M. Stoeckli, P. Chaurand, D. E. Hallahan, and R. M. Caprioli. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nature Medicine*, 7(4):493–496, 2001.

33 E. Sudderth and Michael I. Jordan. Shared segmentation of natural scenes using dependent Pitman-Yor processes. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, Cambridge, MA, 2008. MIT Press.

34 Y. W. Teh, Michael I. Jordan, M. J. Beal, and David M. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2004.

35 Jeramie D. Watrous, Theodore Alexandrov, and Pieter C. Dorrestein. The evolving field of imaging mass spectrometry and its impact on future biological research. *Journal of Mass Spectrometry*, 46:209–222, 2011.