CLUSTERING AND META-CLUSTERING GENE EXPRESSION DATA WITH POSITIVE MATRIX FACTORIZATIONS

Liviu Badea **Doina Ţilivea**

badea@ici.ro

ICI - National Institute for Research and Development in Informatics, Bucharest

Abstract. Although clustering is probably the most frequently used tool for data mining gene expression data, existing clustering approaches face at least one of the following problems in this domain: a huge number of variables (genes) as compared to the number of samples, high noise levels, the inability to naturally deal with overlapping clusters, the instability of the resulting clusters w.r.t. the initialization of the algorithm and/or the difficulty in clustering genes and samples simultaneously. In this paper we show that these problems (except maybe the first) can be elegantly dealt with by using nonnegative matrix factorizations to cluster genes and samples simultaneously while allowing for bicluster overlaps and by employing Positive Tensor Factorization to perform a two-way meta-clustering of the biclusters produced in several different clustering runs (thereby addressing the above-mentioned instability). The application of our approach to a large lung cancer dataset proved computationally tractable and was able to perfectly recover the histological classification of the various cancer subtypes represented in the dataset.

Keywords: bioinformatics, data mining, gene expression data analysis, clustering, meta-clustering.

1. Introduction and motivation

The BIOINFO project aims at developing bioinformatics tools for understanding the mechanisms of complex diseases, such as various types of cancer or type 2 diabetes. The main application domain of this research involves determining diagnostic tools and/or therapeutic targets for these diseases.

The recent advent of high-throughput experimental data, especially in molecular biology and genomics, poses new challenges to existing data mining tools. Measuring the expression levels of virtually every gene of a given organism in a given state has become a routine procedure in many research labs worldwide and has also reached the commercial stage in the last decade. Such gene chips, or *microarrays*, could *in principle* be used to determine the variation in gene expression profiles responsible for a complex disease, such as cancer. However, the large numbers of genes involved (up to a few tens of thousands) compared to the small number of samples (tens to a few hundreds), as well as the large experimental noise levels pose significant challenges to current data mining tools.

Moreover, most currently used clustering algorithms produce *non-overlapping* clusters, which represents a serious limitation in this domain, since a gene is typically involved in several biological processes. In this paper we make a biologically plausible simplifying assumption that the overlap of influences (biological processes) is *additive*

$$X_{sg} = \sum_c X(s,g \mid c) \tag{1}$$

where X_{sg} is the expression level of gene g in data sample s, while $X(s,g \mid c)$ is the expression level of g in s due to biological process c. We also assume that $X(s,g \mid c)$ is multiplicatively decomposable into the expression level A_{sc} of the biological process (cluster) c in sample s and the membership degree S_{cg} of gene g in c:

$$X(s,g \mid c) = A_{sc} \cdot S_{cg} \tag{2}$$

Fuzzy k-means [7] or *Nonnegative Matrix Factorization (NMF)* [4] could be used to produce potentially overlapping clusters, but these approaches are affected by a significant problem: the *instability* of the resulting clusters w.r.t. the initialization of the algorithm. This is not surprising if we adopt a unifying view of clustering as a constrained optimization problem, since the fitness landscape of such a complex problem may involve many different local minima into which the algorithm may get caught when started off from different initial states.

Although such an instability seems hard to avoid, we may be interested in the clusters that keep reappearing in the majority of the runs of the algorithm. This is related to the problem of *combining multiple clustering systems*, which is the unsupervised analog of the classifier combination problem [8], a comparatively simpler problem that has attracted a lot of research in the past decade. Combining clustering results is more complicated than combining classifiers, as it involves solving an additional so-called *cluster correspondence* problem, which amounts to finding the best matches between clusters generated in different runs.

The cluster correspondence problem can also be cast as an unsupervised optimization problem, which can be solved by a *meta-clustering algorithm*. Choosing an appropriate meta-clustering algorithm for dealing with this problem crucially depends on the precise notion of cluster correspondence.

Since a very strict notion of *perfect one-to-one correspondence* between the clusters of each pair of clustering runs may be too tough to be realized in most practical cases, we could look for clusters that are most *similar* (although not necessarily identical) across all runs. This is closest to performing something similar to single-linkage hierarchical clustering on the sets of clusters produced in the various clustering runs, with the additional constraint of allowing in each meta-cluster no more than a single cluster from each individual run. Unfortunately, this constraint will render the meta-clustering algorithm highly unstable. Thus, while trying to address the instability of (object-level) clustering using meta-level clustering, we end up with instability in the meta-clustering algorithm itself. Therefore, a "softer" notion of cluster correspondence is needed.

In this paper, we show that a generalization of NMF called Positive Tensor Factorization (PTF) [6] is precisely the tool needed for meta-clustering "soft", potentially overlapping *biclusters* produced in different clustering runs by fuzzy k-means or NMF. We finally show that the approach is successful at biclustering a large lung cancer gene expression dataset.

2. Generating overlapping clusters with NMF

Combining (1) and (2) leads to a reformulation of our clustering problem as a *nonnegative factorization* of the $n_s \times n_g$ (samples × genes) gene expression matrix X as a product of an $n_s \times n_c$ (samples × clusters) matrix A and an $n_c \times n_g$ (clusters × genes) matrix S:

$$X_{sg} \approx \sum_c A_{sc} \cdot S_{cg} \tag{3}$$

with the additional nonnegativity constraints: $A_{sc} \geq 0, \; S_{cg} \geq 0.$ (4)
(Expression levels and membership degrees cannot be negative.)

More formally, this can be cast as a constrained optimization problem:

$$\min C(A,S) = \frac{1}{2} \| X - A \cdot S \|_F^2 = \frac{1}{2} \sum_{s,g} (X - A \cdot S)_{sg}^2 \tag{5}$$

subject to the nonnegativity constraints (4), and could be solved using Lee and Seung's seminal *Nonnegative Matrix Factorization* (*NMF*) algorithm [4,5], shown below.[1]

NMF(X, A_0, S_0) → (A,S)

$A \leftarrow A_0$, $S \leftarrow S_0$ (typically A_0,S_0 are initialized randomly)

loop

$$S_{cg} \leftarrow S_{cg} \frac{(A^T \cdot X)_{cg}}{(A^T \cdot A \cdot S)_{cg} + \varepsilon}$$

$$A_{sc} \leftarrow A_{sc} \frac{(X \cdot S^T)_{sc}}{(A \cdot S \cdot S^T)_{sc} + \varepsilon}$$

until convergence.

As explained above, such a factorization can be viewed as a "soft" clustering algorithm allowing for *overlapping clusters*, since we may have several significant S_{cg} entries on a given column g of S (so a gene g may "belong" to several clusters c).

Allowing for cluster overlap alleviates but does not completely eliminate the instability of clustering, since the optimization problem (5), (4) is non-convex. In particular, the NMF algorithm produces different factorizations (biclusters) $(A^{(i)},S^{(i)})$ for different initializations, so meta-clustering the resulting "soft" clusters might be needed to obtain a more stable set of clusters. However, using a "hard" *meta*-clustering algorithm would once again entail an unwanted instability.

In this paper we use *Positive Tensor Factorization* (*PTF*) as a "soft" meta-clustering approach able to deal with *biclusters*. This not only alleviates the instability of a "hard" meta-clustering algorithm, but also produces a "base" set of "*bicluster prototypes*", out of which all clusters of all individual runs can be recomposed, despite the fact that they may not correspond to identically reoccurring clusters in all individual runs (see Figure 1).

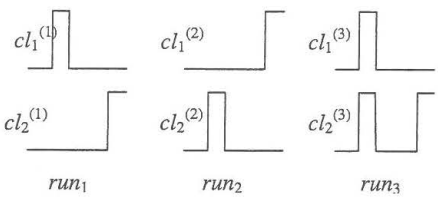

Figure 1. Clusters obtained in different runs are typically combinations of a "base" set of "cluster prototypes" (rather than identical across all runs)

3. Two-way metaclustering with PTF

We use NMF for object-level clustering and PTF for meta-clustering. This unified approach solves in an elegant manner both the clustering and the cluster correspondence problem. More precisely, we first run NMF as object-level clustering r times:

$$X \approx A^{(i)} \cdot S^{(i)} \qquad i = 1,\ldots, r \tag{6}$$

[1] ε is a regularization parameter (a very small positive number).

where X is the gene expression matrix to be factorized (samples × genes), $A^{(i)}$ (samples × clusters) and $S^{(i)}$ (clusters × genes).

To allow the comparison of membership degrees S_{cg} for different clusters c, we scale the rows of $S^{(i)}$ to unit norm by taking advantage of the scaling invariance of the above factorization (6). More precisely:

Proposition. The NMF objective function (5) is invariant under the transformation $A \leftarrow A \cdot D$, $S \leftarrow D^{-1} \cdot S$, where $D = diag(d_1,\ldots,d_{nc})$ is a positive diagonal matrix.

Since a diagonal matrix D operates on the rows of S and on the columns of A, we can scale the rows of S to unit norm by using a diagonal scaling with $d_c = \sqrt{\sum_g S_{cg}^2}$.

Next, we perform *meta-clustering* of the resulting *biclusters* ($A^{(i)}$, $S^{(i)}$). This is in contrast with as far as we know all existing meta-clustering approaches, which take only one dimension into account (either the object- or the sample dimension). Although such *one-way* approaches work well in many cases, they will fail whenever two clusters correspond to very similar sets of genes, while differing along the sample dimension.

In the following, we show that a slight generalization of NMF, namely *Positive Tensor Factorization* (*PTF*) [6] can be successfully used to perform *two-way* meta-clustering, which takes both the gene and the sample dimensions into account.

Naively, one would be tempted to try clustering the biclusters[2] $A_c^{(i)} \cdot S_c^{(i)}$ instead of the gene clusters $S_c^{(i)}$, but this is practically infeasible in most real-life datasets because it involves factorizing a matrix of size $r \cdot n_c \times n_s \cdot n_g$.. On closer inspection, however, it turns out that it is not necessary to construct this full-blown matrix – actually we are searching for a *Positive Tensor Factorization* of this matrix [3]

$$A_{sc}^{(i)} \cdot S_{cg}^{(i)} \approx \sum_{k=1}^{n_c} \alpha_{ck}^{(i)} \cdot \beta_{sk} \cdot \gamma_{kg} \tag{7}$$

The indices in (7) have the following domains: s – samples, g – genes, c – clusters, k – metaclusters. To simplify the notation, we merge the indices i and c into a single index (ic):

$$A_{s(ic)} \cdot S_{(ic)g} \approx \sum_{k=1}^{n_c} \alpha_{(ic)k} \cdot \beta_{sk} \cdot \gamma_{kg} \tag{7'}$$

Note that β and γ are the "unified" versions of $A^{(i)}$ and $S^{(i)}$ respectively. More precisely, the columns β_k of β and the corresponding rows $\gamma_{k\cdot}$ of γ make up a *base set of bicluster prototypes* $\beta_k \cdot \gamma_{k\cdot}$ out of which all biclusters of all individual runs can be recomposed, while α encodes the *(bi)cluster-metacluster correspondence.*

Ideally (in case of a perfect one-to-one correspondence of biclusters across runs), we would expect the rows of α to contain a single significant entry $\alpha_{(ic),m(i,c)}$, so that each bicluster $A_c^{(i)} \cdot S_c^{(i)}$ corresponds to a single bicluster prototype $\beta_{\cdot m(i,c)} \cdot \gamma_{m(i,c)\cdot}$ (where $m(i,c)$ is a function of i and c):

$$A_c^{(i)} \cdot S_c^{(i)} = \alpha_{(ic),m(i,c)} \cdot \beta_{\cdot m(i,c)} \cdot \gamma_{m(i,c)\cdot} \tag{8}$$

---

[2] $A_c^{(i)}$ is the column c of $A^{(i)}$, while $S_c^{(i)}$ is the row c of $S^{(i)}$.

[3] More precisely, we are dealing with the constrained optimization problem

$\min C(\alpha,\beta,\gamma) = \frac{1}{2}\sum_{i,c,s,g}\left(A_{sc}^{(i)}S_{cg}^{(i)} - \sum_{k=1}^{n_c}\alpha_{ck}^{(i)}\beta_{sk}\gamma_{kg}\right)^2$ subject to $\alpha,\beta,\gamma \geq 0$.

Additionally, each meta-cluster m should contain no more than a single bicluster from each individual run, i.e. there should be no significant entries $\alpha_{(ic'),m}$ and $\alpha_{(ic''),m}$ with $c' \neq c''$.

Although it could be easily solved by a hard meta-clustering algorithm, such an ideal cluster correspondence is only very seldom encountered in practice, mainly due to the *instability* of most clustering algorithms.

Thus, instead of such a perfect correspondence (8), we settle for a weaker one (7') in which the rows of α can contain several significant entries, so that all biclusters $A_c^{(i)} \cdot S_c^{(i)}$ are recovered as *combinations* of bicluster prototypes $\beta_k \cdot \gamma_k$.

The nonnegativity constraints of PTF meta-clustering are essential both for allowing the interpretation of $\beta_k \cdot \gamma_k$ as bicluster prototypes as well as for obtaining sparse factorizations. (Experimentally, the rows of α tend to contain typically one or only very few significant entries.)

The factorization (7') can be computed using the following multiplicative update rules (the proofs are straightforward generalizations of those for NMF and can also be found e.g. in [6]):

$$\alpha \leftarrow \alpha * \frac{(A^T \cdot \beta) * (S \cdot \gamma^T)}{\alpha \cdot [(\beta^T \cdot \beta) * (\gamma \cdot \gamma^T)]}$$

$$\alpha \leftarrow \alpha * \frac{(A^T \cdot \beta) * (S \cdot \gamma^T)}{\alpha \cdot [(\beta^T \cdot \beta) * (\gamma \cdot \gamma^T)]} \qquad (9)$$

$$\gamma \leftarrow \gamma * \frac{[\alpha * (A^T \cdot \beta)]^T \cdot S}{[(\alpha^T \cdot \alpha) * (\beta^T \cdot \beta)]^T \cdot \gamma}$$

where '*' and '—' denote element-wise multiplication and division of matrices, while '·' is ordinary matrix multiplication.

After convergence of the PTF update rule, we make the prototype gene clusters directly comparable to each other by normalizing the rows of γ to unit norm ($\|\gamma_k\| = 1$), as well as the columns of α such that $\sum_{i,c} \alpha_{ck}^{(i)} = r$ (r being the number of runs): [4]

$$\gamma_{kg} \mapsto \frac{1}{\|\gamma_k\|} \cdot \gamma_{kg} \qquad \alpha_{(ic)k} \mapsto \frac{r}{\sum_{i',c'} \alpha_{(i'c')k}} \cdot \alpha_{(ic)k} \qquad \beta_{sk} \mapsto \frac{\|\gamma_k\|}{r} \cdot \sum_{ic} \alpha_{(ic)k} \cdot \beta_{sk}$$

and then run NMF initialized with (β, γ) to produce the final factorization $X \approx A \cdot S$ (which can be interpreted as a stable biclustering of X allowing for overlapping clusters).

4. Evaluation on synthetic data

Before addressing real-world gene expression datasets, we evaluated our algorithm on synthetic datasets that match as closely as possible real microarray data. Clusters were modelled using a hidden-variable graphical model as in Figure 2, in which each hidden variable A_c corresponds to the cluster of genes influenced by A_c (clusters can overlap since an observable variable X_g can be influenced by several hidden variables A_c).

---

[4] In order to be able to interpret β and γ as "unified" $A^{(i)}$ and $S^{(i)}$ respectively, we need to have $\sum_c \alpha_{ck}^{(i)} \approx 1$, i.e. $\sum_{i,c} \alpha_{ck}^{(i)} \approx r$, since $X \approx \sum_c A_{sc}^{(i)} \cdot S_{cg}^{(i)} \approx \sum_{k=1}^{n_c} \left(\sum_c \alpha_{ck}^{(i)}\right) \cdot \beta_{sk} \cdot \gamma_{kg}$.

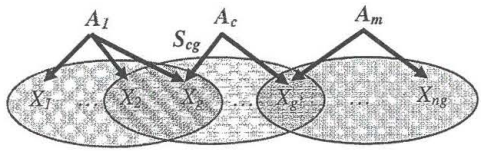

Figure 2. Hidden variable model for generating clusters

Since real-world microarray data are log-normally distributed, we sampled the hidden variables from a $\log_2$-normal distribution with parameters μ=2, σ=0.5, while the influence coefficients S_{cg} between hidden and observable variables were sampled from a uniform distribution over the interval [1,2]. Finally, we added $\log_2$-normally distributed noise ε with parameters μ_{noise}=0, σ_{noise}=0.5. Thus we generated our data using the model $X = A \cdot S + \varepsilon$.

We chose problem dimensions of the order of our real-world application (to be described in the next Section): $n_{samples}$=50, n_{genes}=100, number of genes (respectively samples) per cluster 30 (respectively 15). We compared 4 meta-clustering algorithms (fuzzy k-means, NMF, PTF and the best run[5]) over 10 object-level NMF clustering runs. (Other object level clustering methods perform very poorly and are not shown here). Figures 3-5 below present a comparison of the meta-clustering algorithms w.r.t. the number of clusters (ranging from 2 to 16). The Figures depict average values over 10 separate runs of the whole algorithm (with different randomly generated clusters), as well as the associated SEM bars. Note that although all algorithms produce quite low relative errors $\varepsilon_{rel} = \|X - A \cdot S\| / \|X\|$[6] (under 16%)[7], they behave quite differently when it comes to recovering the original clusters. In a certain way, the match of the recovered clusters with the original ones is more important than the relative error.

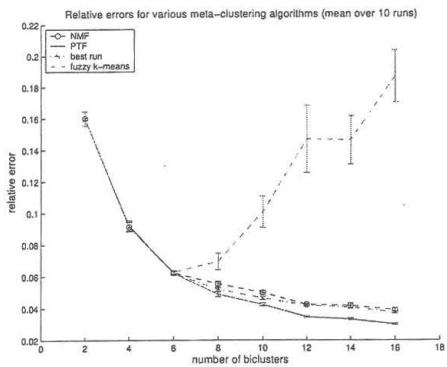

Figure 3. Relative errors versus number of clusters

Defining the *match* between two sets of possibly *overlapping* clusters is nontrivial. For each cluster C_1 from clustering 1, we determine the single cluster C_2 from clustering 2 into which it is best included, i.e. the one with the largest $|C_1 \cap C_2| / |C_1|$. We proceed analogously

[5] i.e. the one with the smallest relative error.
[6] (A_f, S_f) are the factorizations output by the algorithms.
[7] Except for fuzzy k-means which misbehaves for large numbers of clusters.

for the clusters C_2 from clustering 2. Then, for each cluster C_1 (from clustering 1), we determine its match $|C_1 \cap C_2|/|C_1 \cup C_2|$ with the *union* C_2 of clusters from clustering 2, for which C_1 is the best including cluster (as determined in the previous step). Similarly, we determine matches for clusters C_2 from clustering 2. The average match of the two clusterings is then the mean of all these matches (for all C_1 and all C_2).

Figure 4 shows that PTF consistently outperforms the other meta-clustering algorithms in terms of recovering the original clusters. Note that since clusters were generated randomly, their overlap increases with their number, so it is increasingly difficult for the meta-clustering algorithm to discern between them, leading to a decreasing match. This can be directly seen in Figure 5, where we depict both the cluster overlaps (in the initial data) and the matches of the recovered clusters with the original ones. The inverse correlation between bicluster overlap and matches is obvious (Pearson correlation coefficient -0.92).

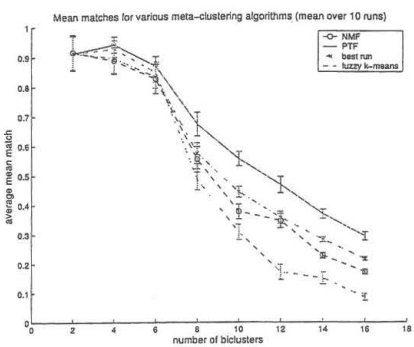

Figure 4. Mean match versus number of clusters

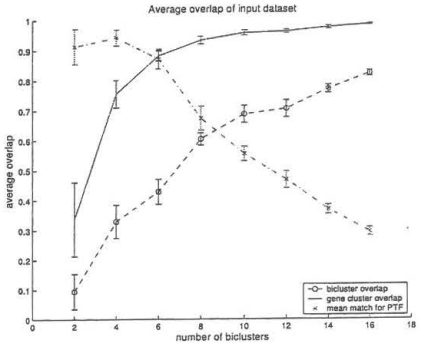

Figure 5. Overlaps and matches are inversely correlated

Among all *object-level* clustering algorithms tried (k-means, fuzzy k-means and NMF), only NMF behaves consistently well. The conceptual elegance of the combination of NMF as object-level clustering and PTF as meta-clustering thus pays off in terms of performance.

5. Metaclustering a lung cancer gene expression dataset

In the following we show that metaclustering is successful at biclustering a large lung cancer dataset from the Meyerson lab [10].

Using HG-U95Av2 Affymetrix oligonucleotide microarrays, Bhattacharjee et al. [10] have measured mRNA expression levels of 12600 transcript sequences (genes) in 186 lung tumor samples (139 adenocarcinomas, 21 squamous cell lung carcinomas, 6 small cell lung cancers, 20 pulmonary carcinoids) and 17 normal lung samples (203 samples in total).

Since the raw CEL files (originating from the scanner software) were presumably processed by the authors with older Affymetrix MAS4 software, certain gene expression value estimations are negative[8]. Therefore, we applied separate additive corrections for genes having negative values so that all gene expression values become positive. (We avoided a global scaling of the samples since various forms of cancer may affect a large fraction of the genome.)

[8] This is due to MAS4 using a $PM - MM$ model, where PM is the expression level of perfect match probes, while MM that of mismatch probes.

Since the raw CEL files (originating from the scanner software) were presumably processed by the authors with older Affymetrix MAS4 software, certain gene expression value estimations are negative[9]. Therefore, we applied separate additive corrections for genes having negative values so that all gene expression values become positive. (We avoided a global scaling of the samples since various forms of cancer may affect a large fraction of the genome.)

For testing our metaclustering algorithm, we first selected a subset of genes that are differentially expressed between the classes. ANOVA, pairwise t-tests or SAM [11] could have been used for this purpose, but we preferred the following SNR measure, since it discourages large intra-class STD in both classes:

$$SNR_{class} = \frac{\mu_{class} - \mu_{normal}}{\sigma_{class} + \sigma_{normal}}.$$

More precisely, we selected the genes with an average expression level over 100 [10] and having $|SNR_{class}| > 2$ for at least one of the classes (small cell, squamous or carcinoid). There were 251 such genes.

Since adenocarcinoma is a very heterogeneous disease, whose subclasses are poorly understood at the molecular level, we discarded the adeno samples from the dataset and used the histological classification of samples provided in the supplementary material to the original paper [1010] as a gold standard for the evaluation of the biclustering results.

To eliminate the bias towards genes with high expression values, the resulting restricted gene expression matrix was then normalized by separate scalings of the genes such that their norms (uncentred STDs) become equal.

Although nonnegative factorizations have the advantage of obtaining sparse and easily interpretable[11] decompositions, they cannot directly account for gene down-regulation. To deal with gene down-regulation in the context of NMF, we extended the gene expression matrix with new "down-regulated genes" g' associated to the original genes g as follows:

$$g' = \mathrm{pos}(mean(g_{normal}) - g)$$

where $mean(g_{normal})$ is the average of the gene over the *normal* samples, while pos(·) is the Heaviside step function.[12]

We then used our metaclustering algorithm to factorize the extended gene expression matrix as follows (with n_c=4 and running PTF over 20 NMF runs):

$$X_{sg} \approx \sum_c A_{sc} \cdot S_{cg}$$

(The matrix X has 64 rows (samples) and 2×251=502 columns (extended genes).)

The following Figure 6 shows the resulting sample matrix *A*. Note that the algorithm has recovered the sample clusters with high accuracy (as can be seen in Figure 6).

The relative error of the decomposition is $\varepsilon = \frac{\|X - AS\|_F}{\|X\|_F} = 0.2722$, while the relative errors of the 20 individual runs are slightly higher.

---

[9] This is due to MAS4 using a $PM - MM$ model, where PM is the expression level of perfect match probes, while MM that of mismatch probes.

[10] For Affymetrix chips, expression levels below 100 are considered to be too low to be reliable.

[11] Since no complex cancellations of positive and negative terms are allowed.

[12] $\mathrm{pos}(x) = x$ if $x > 0$ and 0 otherwise.

Cluster membership degrees S_{cg} were considered significant if they were larger than the threshold $\theta_s = 1/\sqrt{n} = 0.125$.

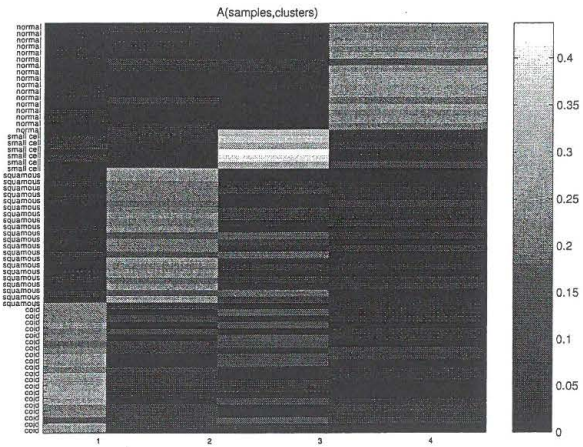

Figure 6. The sample clusters

Note that the overlap between the small cell and carcinoid sample clusters[13] has a biological interpretation: both contain samples of tumors of neuroendocrine type. The low mixing coefficients indicate however that carcinoids are highly divergent from the malignant small cell tumors.

We also looked in detail at some known marker genes. For example, the known small cell marker *ASCL1* (achaete scute 1) is *specific* to the small cell cluster, while *KRT5* (keratin 5) is specific to the squamous cluster.

On the other hand, known proliferative markers like *PCNA* (proliferating cell nuclear antigen), *MCM2* and *MCM6* are *common* to small cell *and* squamous clusters, as expected.

Overall, our metaclustering algorithm proved quite robust at rediscovering the known histological classification of the various lung cancer types in the Meyerson dataset.

6. Related work and conclusions

In this paper we show that nonnegative decompositions such as NMF and PTF can be combined in a non-trivial way to obtain an improved meta-clustering algorithm for *gene expression data*. The approach deals with *overlapping clusters* and alleviates the annoying *instability* of currently used algorithms by using an advanced two-way meta-clustering technique based on *tensor* (rather than matrix) factorizations.

The main contribution of this paper consists however in applying our PTF metaclustering approach to the Meyerson lab lung cancer dataset [10], for which the algorithm was able to perfectly recover the known histological classification of the various lung cancer types represented in the dataset.

Related work by Bradley and Fayyad [1] uses *k-means* for meta-clustering a number of *k-means* runs on subsamples of the data for initializing a final k-means run. The approach was aimed at handling large datasets rather than improving the stability of clustering. So it is

[13] Columns 3 and 1 of A in Figure 6.

not surprising that the use of a "hard" clustering approach like k-means in domains featuring *overlapping bi*clusters produces dramatically less accurate results than our approach using PTF and NMF.

Our approach is also significantly different from other biclustering approaches, such as Cheng and Church''s biclustering [9], which is based on a simpler additive model that is not scale invariant (and thus problematic in the case of gene expression data).

Acknowledgements

This work has been supported by the BIOINFO project (CEEX contract 32/2005) within the Romanian Research of Excellence Programme.

References

1. P.S. Bradley, U.M. Fayyad Refining Initial Points for K-Means Clustering, Proc. ICML-98, pp. 91-99.
2. M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein. Cluster analysis and display of genome-wide expression patterns, PNAS 95, 14863-8
3. P.O. Hoyer Non-negative sparse coding. Neural Networks for Signal Processing XII, 557-565, Martigny, 2002.
4. Lee D.D., H.S. Seung. Learning the parts of objects by non-negative matrix factorization. Nature, vol. 401, no. 6755, pp. 788-791, 1999.
5. Lee D.D., H.S. Seung. Algorithms for non-negative matrix factorization. Proc. NIPS*2000, MIT Press, 2001.
6. M. Welling, M. Weber Positive tensor factorization. Pattern Recognition Letters 22(12): 1255-1261 (2001).
7. J.C. Bezdek Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, 1981.
8. E. Bauer, R. Kohavi An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine Learning 36 (1999) 105-139.
9. Y. Cheng, G. Church Biclustering of expression data. Proc. ISMB-2000, 93-103.
10. Bhattacharjee et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA. 2001 Nov. 20;98(24):13790-5.
11. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. PNAS 2001 98(9):5116-21.