A Probability-Based Combination Method for Unsupervised Clustering with Application to Blind Source Separation

Julian Becker, Martin Spiertz, and Volker Gnann

Institut für Nachrichtentechnik, RWTH Aachen University,
52056 Aachen, Germany
{becker,gnann}@ient.rwth-aachen.de
http://www.ient.rwth-aachen.de

Abstract. Unsupervised clustering algorithms can be combined to improve the robustness and the quality of the results, e.g. in blind source separation. Before combining the results of these clustering methods the corresponding clusters have to be aligned, but usually it is not known which clusters of the employed methods correspond to each other. In this paper, we present a method to avoid this correspondence problem using probability theory. We also present an application of our method in blind source separation. Our approach is better expandable than other state-of-the-art separation algorithms while leading to slightly better results.

1 Introduction

The idea of combining the results of multiple clustering methods has been presented in [1],[2],[3]. For clustering of data, a number of approaches may be applied, usually leading to different results. The intention of combining multiple clustering approaches is to improve the results by using the strengths of all the methods. Unfortunately, in blind clustering methods, the correspondences of the clusters are unknown. When combining methods this correspondence problem has to be solved, see [3]. Fred and Jain propose to use a measure of similarity between patterns [2] to circumvent this problem. We propose a similiar method, extending the approach by using probabilities, which opens another way of clustering the combined results.
An application where multiple clustering methods can be used is blind source separation (BSS). BSS tries to separate the original sources out of a mixed audio signal and can be used as a preprocessing step for many audio processing tasks such as remixing, instrument recognition, or automatic music transcription. Many state-of-the-art algorithms use non-negative tensor factorization (NTF) or non-negative matrix factorization (NMF) to factorize single notes out of the mixture. While [4] and [5] propose extensions to the NTF to factorize complete melodies, [6] presents an approach, where the single notes are being clustered to melodies. In [7], different clustering methods are proposed, one using spectral and the other one temporal features.

In this paper we propose an approach for combining multiple clustering methods, which is presented in Section 2. In Section 3 we use our algorithm to combine the different clustering methods for BSS as proposed in [7] and analyze the performance in comparison to other approaches. Finally in Section 5, a conclusion is given.

2 The Proposed Combination Algorithm

In this section we present the proposed combination strategy for unsupervised clustering methods. We assume a testset of I data items that have to be grouped into C clusters. In the following items will be named i_m with $1 \leq m \leq I$ and clusters c with $1 \leq c \leq C$. We furthermore assume a number of V different clustering methods, which all return probabilities ${}^{v}p_c(i_m)$ that item i_m belongs to cluster c with $1 \leq v \leq V$. Thus, every method returns a matrix of size $I \times C$. Combining these matrices is not possible because it is not known which clusters of the different methods correspond to each other. So, before combining the matrices it is first necessary to estimate the correspondences of the clusters and to align the columns. This step may induce errors if the correspondences are not estimated correctly. This issue motivates our proposed algorithm.

2.1 The Basic Idea

Instead of evaluating the probabilitiy $p_c(i_m)$ that item i_m belongs to cluster c, we propose to calculate the probability $p(i_m, i_n)$ that the items i_m and i_n belong to the same cluster. This means, for every clustering method v we calculate a matrix

$$\mathbf{Q}_v = \begin{pmatrix} 1 & {}^{v}p(i_1, i_2) & \cdots & {}^{v}p(i_1, i_I) \\ {}^{v}p(i_2, i_1) & 1 & \cdots & {}^{v}p(i_2, i_I) \\ \vdots & \vdots & \ddots & \vdots \\ {}^{v}p(i_I, i_1) & {}^{v}p(i_I, i_2) & \cdots & 1 \end{pmatrix} \tag{1}$$

where the entries ${}^{v}p(i_m, i_n)$ are calculated as

$${}^{v}p(i_m, i_n) = \sum_{k=1}^{C} {}^{v}p_k(i_m) \cdot {}^{v}p_k(i_n) \tag{2}$$

for each clustering method v, leading to V matrices $\mathbf{Q}_v$ of size $I \times I$.
These matrices $\mathbf{Q}_v$ can now be combined without having to be aligned. One possibility to combine the matrices is taking the mean values of the entries ${}^{v}p(i_m, i_n)$ over all v. This leads to a matrix $\mathbf{Q}_{\text{av}}$ with the average probabilities $p_{\text{av}}(i_m, i_n)$

$$\mathbf{Q}_{\text{av}} = \begin{pmatrix} 1 & p_{\text{av}}(i_1, i_2) & \cdots & p_{\text{av}}(i_1, i_I) \\ p_{\text{av}}(i_2, i_1) & 1 & \cdots & p_{\text{av}}(i_2, i_I) \\ \vdots & \vdots & \ddots & \vdots \\ p_{\text{av}}(i_I, i_1) & p_{\text{av}}(i_I, i_2) & \cdots & 1 \end{pmatrix} \tag{3}$$

where the entries $p_{\text{av}}(i_m, i_n)$ are calculated as

$$p_{\text{av}}(i_m, i_n) = \frac{\sum_{v=1}^{V} {}^{v}p(i_m, i_n)}{V}. \tag{4}$$

The indices m and n corresponding to the maximum value in $\mathbf{Q}_{\text{av}}$ denote the items that are most probable to belong to the same cluster.
Other combinations of the matrices are possible. For example instead of taking the mean value, the different methods could be weighted, for example depending on how good they perform. A weighted combination will be used in Section 3.2.

2.2 The Clustering Algorithm

The matrix $\mathbf{Q}_{\text{av}}$ can now be used for clustering. In our clustering algorithm, items will be iteratively grouped together. These groups of items will be named ${}^{q}Z_r$ where q is the current iteration step and $1 \leq r \leq R$ indicates all existing groups. Every group contains at least one item.
Groups can also be interpreted as events in a probability meaning. Every group represents the event, that the items in this group belong to the same cluster.
We define the following notations:

Term	**Meaning**
$p({}^{q}Z_r)$	Probability that all items that are grouped together in group ${}^{q}Z_r$ belong to the same cluster
$p(\{{}^{q}Z_r, {}^{q}Z_s\})$	Probability that all items that are grouped together in the groups ${}^{q}Z_r$ and ${}^{q}Z_s$ belong to the same cluster
$p({}^{q}Z_r \cap {}^{q}Z_s)$	Probability that the events ${}^{q}Z_r$ and ${}^{q}Z_s$ both occur
${}^{q}Z$	Unites all events ${}^{q}Z_1, {}^{q}Z_2, \ldots, {}^{q}Z_R$. This means the notation $p({}^{q}Z)$ describes the same probability as $p({}^{q}Z_1 \cap {}^{q}Z_2 \cap \ldots \cap {}^{q}Z_R)$
${}^{q+1}Z_r = \{{}^{q}Z_s, {}^{q}Z_t\}$	Indicates, that in the iteration step $q+1$, all items that were grouped together in ${}^{q}Z_s$ and ${}^{q}Z_t$ are merged in group ${}^{q+1}Z_r$

For our clustering algorithm we will need to calculate a matrix similiar to the matrix $\mathbf{Q}_{\text{av}}$ in Eq. (3) in every iteration step. This matrix is denoted ${}^{q}\tilde{\mathbf{Q}}_{\text{av}}$ and will be called probability matrix in the following. The entries at index m, n are now defined as

$$p_{\text{av}}(\{{}^{q}Z_m, {}^{q}Z_n\} | {}^{q}Z) = \frac{\sum_{v=1}^{V} {}^{v}p(\{{}^{q}Z_m, {}^{q}Z_n\} | {}^{q}Z)}{V}. \tag{5}$$

We assume that the events ${}^{q}Z_r$ and ${}^{q}Z_s$ are independent, if $r \neq s$. In this case the probabilitiy $p({}^{q}Z_r \cap {}^{q}Z_s)$ reduces to

$$p({}^{q}Z_r \cap {}^{q}Z_s) = p({}^{q}Z_r) \cdot p({}^{q}Z_s). \tag{6}$$

Considering the fact that ${}^{q}Z_s$ is a subset of $\{{}^{q}Z_r, {}^{q}Z_s\}$ it is obvious that

$$p(\{{}^{q}Z_r, {}^{q}Z_s\} \cap {}^{q}Z_s) = p(\{{}^{q}Z_r, {}^{q}Z_s\}). \tag{7}$$

Using the definition of conditional probability, the probability ${}^{v}p(\{{}^{q}Z_m, {}^{q}Z_n\}|{}^{q}Z)$ in Eq. (5) can be written as

$${}^{v}p(\{{}^{q}Z_m, {}^{q}Z_n\}|{}^{q}Z) = \frac{{}^{v}p(\{{}^{q}Z_m, {}^{q}Z_n\} \cap {}^{q}Z)}{{}^{v}p({}^{q}Z)}. \tag{8}$$

With Equations 6 and 7 this term reduces to

$${}^{v}p(\{{}^{q}Z_m, {}^{q}Z_n\}|{}^{q}Z) = \frac{{}^{v}p(\{{}^{q}Z_m, {}^{q}Z_n\})}{{}^{v}p({}^{q}Z_m) \cdot {}^{v}p({}^{q}Z_n)}. \tag{9}$$

This group-based definition of ${}^{q}\tilde{\mathbf{Q}}_{\mathrm{av}}$ allows us to group items together iteratively. For the special case that every group ${}^{q}Z_r$ contains exactly one item, the matrix ${}^{q}\tilde{\mathbf{Q}}_{\mathrm{av}}$ is identical to $\mathbf{Q}_{\mathrm{av}}$ in Eq. (3).

We suggest the following iterative algorithm for combined clustering:

```
Initialize ${}^{0}Z_r = \{i_r\} \, \forall \, r = 1, 2, \ldots, I$
$q = 0$, $q_{\max} = I - C$
while $q < q_{\max}$ do
    Calculate ${}^{q}\mathbf{Q}_{\mathrm{av}}$ (Eq. (3) and (5))
    $m, n = \underset{\tilde{m},\tilde{n}}{\mathrm{argmax}} \, {}^{q}\mathbf{Q}_{\mathrm{av}}(\tilde{m}, \tilde{n}), \ \tilde{m} < \tilde{n}$
    ${}^{q+1}Z_r = \begin{cases} {}^{q}Z_r & \text{for } r < m \\ \{{}^{q}Z_m, {}^{q}Z_n\} & \text{for } r = m \\ {}^{q}Z_r & \text{for } m < r < n \\ {}^{q}Z_{r+1} & \text{for } r \geq n \end{cases}$
    $q = q + 1$
end while
Return ${}^{q_{\max}}Z_r$
```

The C remaining groups ${}^{q_{\max}}Z_r$ are the result of the clustering. Every group can be interpreted as one cluster. All items that belong to this group are assigned to this cluster.

3 Application to Blind Source Separation

In the following the algorithm is applied to BSS. In [7], an approach for BSS was presented which uses two clustering methods. The methods use spectral and temporal information, respectively. The combination of both methods was done by hard-decision. However, it seems reasonable to assume, that even if one of the methods is more reliable, the other method still contains information that could improve the clustering. Hence a soft-decision combination could improve the results.

3.1 Hard-Decision Approach

More detailed information about the hard-decision approach can be found in [7] and [6].
In the following we assume $x(n)$ to be an additive mixture of M monaural sources $s_m(n)$, with n being the time index. In the following we present the signal flow of the algorithm.

- First, the short-time Fourier transform (STFT) of $x(n)$ is taken. The resulting complex-valued spectrogram $\underline{\mathbf{X}}$ is of size $K \times T$ with frequency-bins $1 \leq k \leq K$ and time-bins $1 \leq t \leq T$. In the following only the absolute values, $\mathbf{X} = |\underline{\mathbf{X}}|$, are used.
- In the next step, $\mathbf{X}$ is factorized by the NMF. This results in a separation of I sound events. The NMF outputs the two matrices $\mathbf{B}$ of size $K \times I$ and $\mathbf{G}$ of size $T \times I$. These matrices approximate $\mathbf{X}$ by
$$\mathbf{X}(k,t) \approx \sum_{j=1}^{I} \mathbf{B}(k,j)\mathbf{G}(t,j). \tag{10}$$
The j-th column of $\mathbf{B}$ corresponds to the spectrum and the j-th column of $\mathbf{G}$ represents the temporal envelope of sound event σ_j. For multichannel signals, the NTF can be used instead of the NMF.
- Signal synthesis is done as described in [6]. The spectrogram $\underline{\mathbf{Y}}_j(k,t)$ corresponding to the estimated time domain signal $y_j(n)$ of sound event σ_j is calculated as:
$$\underline{\mathbf{Y}}_j(k,t) = \underline{\mathbf{X}}(k,t) \cdot \frac{\mathbf{B}(k,j)\mathbf{G}(t,j)}{\sum_{z=1}^{I} \mathbf{B}(k,z)\mathbf{G}(t,z)}. \tag{11}$$
The output signals $y_j(n)$ are estimated by applying the inverse STFT to $\underline{\mathbf{Y}}_j$.
- The I sound events are clustered into M clusters. A vector $\mathbf{a}$ with I elements is defined, with $1 \leq \mathbf{a}(j) \leq M$, $\mathbf{a}(j) \in \mathbb{N}$. The entries $\mathbf{a}(j)$ of this vector specify the cluster, to which cluster the sound event σ_j is assigned. Clustering is done using the NMF as proposed in [6].
Features $\mathbf{F_B}$ and $\mathbf{F_G}$ are calculated from the matrices $\mathbf{B}$ and $\mathbf{G}$ using the source-filter model theory for frequency and time domain [7]. The features are independently factorized by an NMF, which gives an approximation
$$\mathbf{F}_{\{\mathbf{B}|\mathbf{G}\}}(k,j) \approx \sum_{m=1}^{M} \mathbf{W}_{\{\mathbf{B}|\mathbf{G}\}}(k,m)\mathbf{V}_{\{\mathbf{B}|\mathbf{G}\}}(j,m). \tag{12}$$
The index $\{\mathbf{B}|\mathbf{G}\}$ denotes, that either the matrices calculated from $\mathbf{B}$ or from $\mathbf{G}$ are used. While the m-th column of $\mathbf{W}$ corresponds to the m-th cluster center, the m-th column of $\mathbf{V}$ corresponds to the m-th connectivity values. Therefore the clustering vector $\mathbf{a}$ is defined as
$$\mathbf{a}(j) = \underset{m}{\mathrm{argmax}} \mathbf{V}(j,m). \tag{13}$$

Hence, we get one vector $\mathbf{a}_{\mathbf{B}}(j)$ from the spectral clustering and one vector $\mathbf{a}_{\mathbf{G}}(j)$ from the temporal clustering.

- The decision, which clustering vector to be used is based on the *number of note instances* μ_m of the mixture. This value μ_m is calculated for each column of $\mathbf{G}$ by subtracting the mean value of the corresponding column and counting the zero crossings from negative to positive values. The final value μ_{av} is estimated as the mean value over all μ_m. The clustering vector $\mathbf{a}_{final}(j)$ that is applied for the final clustering is $\mathbf{a}_{\mathbf{B}}(j)$ if $\mu_{\text{av}} \leq \vartheta$, or else $\mathbf{a}_{\mathbf{G}}(j)$, with ϑ being a predefined threshold. In [7], a value of $\vartheta = 1.6$ is proposed.

3.2 Extension to Soft-Decision

Instead of using a hard-decision we propose a soft-decision combination of the clustering methods, using the algorithm presented in Section 2. The correspondences between the given problem and the proposed algorithm are as follows:

Clusters The clusters for the algorithm are the different estimated sources of the mixture signal. The number of clusters C in the algorithm is therefore M.

Items The items i_m of the algorithm are the I separated sound events σ_j.

Methods The clustering methods that have to be combined are the spectral and the temporal clustering methods.

Probabilities Our combination algorithm requires probabilities instead of hard mappings. We define

$$p_m(\sigma_j) = \frac{\mathbf{V}(j,m)}{\sum_{k=1}^{M} \mathbf{V}(j,k)}. \tag{14}$$

This definition leads to a matrix of probabilities of size $I \times M$ for every clustering method, which can be used as input for the proposed algorithm.

Weightings Instead of the hard-decision a soft-decision is made by weighting the probability matrices of the different methods before combining them. The probability matrix corresponding to the spectral clustering is weigthed with $w_{\mathbf{B}}$ with

$$w_{\mathbf{B}} = \begin{cases} 1 & \text{if } \mu_{\text{av}} \leq b_l \\ \frac{b_u - \mu_{\text{av}}}{b_u - b_l} & \text{if } b_l < \mu_{\text{av}} \leq b_u \\ 0 & \text{if } \mu_{\text{av}} > b_u \end{cases}, \tag{15}$$

where b_l and b_u denote a lower and an upper bound. These parameters have to be determined by experiment. The probability matrix corresponding to the temporal clustering is weigthed with $1 - w_{\mathbf{B}}$. For the special case $b_l = b_u$ this transforms to the hard-decision criterion with threshold b_l.

4 Experimental Results

We compare our soft-decision approach with the hard-decision approach in [7]. We also compare our results with the results of [4].

For performance measurement we use the measures SDR, SIR and SAR as proposed in [8]. The test set 1 that we use for comparison with [7] is a set of 1770 two-source mixtures, mixed from 60 monaural recordings, which is identical to test set $\mathcal{A}$ in [7]. Test set 2 are the 25 mixtures used in [4]. This dataset mainly contains very harmonic mixtures.

For fair comparison, we use exactly the same parameters for our algorithm as are used in [7]. In [7] a value of $\vartheta = 1.6$ is used as threshold for the hard-decision. Therefore we chose the values for the upper and the lower bound of the weights for the soft-decision symmetrical around this value. Experiments show, that values of $b_l = 0.8$ and $b_u = 2.4$ lead to good results.

The mean values over SDR, SIR and SAR for test set 1 are shown in Table 1. It can be observed that the soft-decision approach performs slightly better than the approach of [7] for all of the three measures.

The mean values over SDR, SIR and SAR for test set 2 are shown in Table 2. We compare our results with the results of the hard-decision approach [7] and with the results of [4]. Compared to the algorithm in [4], our algorithm leads to lower distortion by artifacts (SAR) but to higher distortion by interferences (SIR). Compared to the hard-decision approach [7] our algorithm leads to slightly better results for all of the three measures. In [7] it is shown that for test set 2, spectral clustering leads to much better results than temporal clustering, which can be explained by the high harmonicity of the sources. However, our results show, that even for such harmonic mixtures the results can be improved by also using temporal information.

Test set 1	SDR	SIR	SAR
hard-decision [7]	7.20	12.92	13.62
soft-decision	7.23	13.07	13.83

Table 1. Results for SDR, SIR and SAR in dB for test set 1.

Test set 2	SDR	SIR	SAR
[4]	9.01	24.91	9.52
hard-decision [7]	9.80	15.05	15.91
soft-decision	9.91	15.21	16.27

Table 2. Results for SDR, SIR and SAR in dB for test set 2.

It should be noted that besides the slightly better results compared to the hard-decision approach, presented in Table 1 and Table 2, the proposed combination algorithm holds the advantage of beeing easily expandable by appending other matrices of clustering results to the input. It can be assumed that by including more methods, the results could be improved further. This possibility of using more clustering methods is not given in the hard-decision approach.

5 Conclusion

In this paper we present a new way of combining different clustering methods based on probability theory. We calculate the probabilities that different items belong to the same cluster, which makes it possible to combine different methods without having to solve the correspondence problem. We introduce a method of clustering the combined values by iteratively grouping together items that most probably belong to the same cluster.

We use the presented approach to extend the BSS-algorithm proposed in [7] by using a soft-decision combination. We show that this extension leads to slightly better separation results. Furthermore, our approach has the advantage of being easily expandable, using more clustering methods.

References

1. A. Strehl and J. Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3:583–617, March 2003.
2. C.Boulis and M. Ostendorf. Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, 2005.
3. A.L.N. Fred and A.K. Jain. Combining multiple clustering systems. *Knowledge Discovery in Databases: PKDD 2004*, volume 3202 of *Lecture Notes in Computer Science*, pages 63–74. 2004.
4. D. FitzGerald, M. Cranitch, and E. Coyle. Extended nonnegative tensor factorisation models for musical sound source separation. *Computational Intelligence and Neuroscience*, 2008.
5. A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing* 18(3):550–563, 2010. `http://www.irisa.fr/metiss/ozerov/demos.html`.
6. M. Spiertz and V. Gnann. Source-filter based clustering for monaural blind source separation. *Proc of International Conference on Digital Audio Effects DAFx*, Como, Italy, 2009.
7. M. Spiertz and V. Gnann. Note clustering based on 2d source-filter modeling for underdetermined blind source separation. *Proceedings of the AES 42nd International Conference on Semantic Audio*, Ilmenau, Germany, July 2011.
8. E. Vincent, R. Gribonval, and C. Fevotte. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing* 14(4):1462–1469, 2006.