Experimental investigation of linear mixing in real world datasets

Jie Ouyang
Thomas Trappenberg
Andrew Back

Technical Report CS-2004-05

May 31,2004

Faculty of Computer Science
6050 University Ave., Halifax, Nova Scotia, B3H 1W5, Canada

Experimental investigation of linear mixing in real world datasets

Jie Ouyang, Thomas Trappenberg
Faculty of Computer Science, Dalhousie University, Halifax, NS, B3H 1W5,Canada
{ouyang,tt}@cs.dal.ca
Andrew D. Back
Windale Technologies, Brisbane, QLD 4075, Australia
andrew@andrewback.com

May 31, 2004

Abstract

It is well acknowledged in the data mining community that feature values, which become the input variables for modeling the system, are often statistically dependent. In this paper we attempt to quantify the dependencies by assuming a linear mixing model and using an independent component analysis (ICA) to estimate the mixing matrix. The major difficulty in quantifying the mixing strength comes thereby from the fact that ICA algorithms give estimations of a mixing matrix only up to row permutations and scalar factors of the mixing matrix. In this paper we propose several measures of the mixing strength that are either appropriate estimates or lower bounds of the true linear mixing strength. These measures are tested on generic data and on 30 datasets from standard machine learning repositories. The experimental results not only indicate that statistical mixtures between input variables exist in real world problems, but most of them are strong.

1 Introduction

Systems that we want to model for data mining purposes are often characterized experimentally by listing feature values in datasets. For example, a patient in a hospital might be captured on file by his identity number, gender, age, weight, blood pressure, blood glucose levels, and various other measurements that might be necessary to monitor his health state. It is clear that there could be dependencies between those feature values. For example, overweight increases the risk of type 2 diabetes, and resulting health problems from diabetes might increase the likelihood of high blood pressure. Therefore, each feature values does not add independent information, which should be taken into account in data mining techniques. Indeed, it has been shown that combinations of input variable, which result in a more suitable representation for the data mining algorithm, can drastically advance applications of knowledge discovery and data mining [1, 2, 3].

A related area where mixing of feature values should be considered is input variable selection [4]. This topic has received renewed interest due to the increasing size of datasets that are available in many application areas [5]. We have previously shown that an independent component analysis (ICA) preprocessing step can be beneficial for input variable selection [6]. For example, the number of necessary input variables can be overestimated by some variable selection schemes in the case when the input signals are mixtures of source signals of which only some determine the output signal. This papers follows up on this idea by attempting to verify if such mixing is common in real world datasets.

While it is well acknowledged that data dependencies between features are common in many applications, no attempt has been made to our knowledge to quantify the dependencies in real world datasets. We believe that a better characterization and quantification of dependencies between feature values can help in the further improvement of data mining techniques. In this paper we report on our attempt to quantify how *strong* the mixing between feature values in real world datasets is. The reason we concentrate on the strength of mixing is that preprocessing of data with ICA can be particularly beneficial in data sets with considerable mixing. Quantification of the mixing is difficult without specific knowledge of the nature of mixing itself. In this paper we assume a linear mixing model so that we can apply standard ICA algorithms.

In order to estimate the mixing in real world datasets we apply the FastICA algorithms [7] on the feature values from randomly selected datasets and attempt to quantify the mixing strength based on the sum of off-diagonal elements of the (estimated) mixing matrix. A problem with this approach is that ICA algorithms can not determine the order of the source signals so that the estimated mixing matrix is only an estimate of a linear mixing matrix up to permutations of columns. In other words, the statistical dependencies determined by ICA can be caused by different mixing models. However, as the major aim in this study is to see if strong mixing between features can be established in some databases, we concentrate on the conservative estimate by finding the minimal possible mixing strength in a linear mixing model by considering all permutations of the estimated mixing matrix.

A further complication of searching for the matrix with minimal mixing strength in all possible column-permutations of the estimated mixing matrix is the computational complexity as the number of permutations is $n!$, where n is the number of features in the dataset. One of the main contributions of this paper is to propose a new measure called E_1^{C} which we show to be a strict lower bound on the minimal mixing strength of all column-permutations of the estimated mixing matrix. A major advantage of this measure is that it has a greatly reduced computational complexity of $O(n^2)$ instead of $O(n!)$. As an example, for a data set of only 12 variables, this amounts to an improvement in efficiency of 3,326,400 $\times$ fewer operations. In addition, we introduce a measure E_1^{N}, closely related to the quantity E_1 proposed by Amari et al. [8], which is not a lower bound of the minimal mixing strength but can better estimate the more likely mixing strength. Both measures, E_1^{C} and E_1^{N}, are applied to 30 datasets from which we found considerable evidence that strong mixing is present in most of those datasets.

2 Background and Problems

2.1 Using ICA to estimate a linear mixing matrix

Independent component analysis is a technique to recover statistically independent source signals from the measurement of their mixtures (see [9] for a review). The basic models assumes a linear mixing of source signals,

$$\mathbf{x} = \mathbf{M}\mathbf{s}, \tag{1}$$

where $\mathbf{M}$ is an generally unknown, instantaneous *mixing* matrix, $\mathbf{s}$ is a vector of unobserved independent source signals, and $\mathbf{x}$ is a vector of observed dependent signals. Several algorithms [10, 11, 8, 7, 9], have been developed to estimate the source signals based on the minimization of statistical dependencies between the observed signals,

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}, \tag{2}$$

where $\hat{\mathbf{s}}$ is a vector of estimated independent signals and $\mathbf{W}$ is the *de-mixing* matrix. For the following discussions it is important to note that the order and overall strength of the estimated source signals can not be estimated by ICA. Thus, the inverse of the de-mixing matrix $\mathbf{A} = \mathbf{W}^{-1}$ is only an estimate of the mixing $\mathbf{M}$ up to a permutation of columns and some scale factor for each row.

Typical applications of ICA are mostly interested in recovering source signals, but we are here using ICA for estimating the mixing matrix. More specify, we are interested in using this estimate to quantify the strength of the mixing which requires a definition of the measure where the problem of column permutations and scale factors have to be taken into account. In the remainder of this section we propose several measures and outline their specific limitations.

2.2 Definition of mixing strength: E_1^{M}

A mixing matrix which is diagonal correspond to no mixing between source signals. With a definition of a strength measure we want to capture the off-diagonality of the matrix. We thus define the *mixing strength* basically as the sum of the absolutes of the off-diagonal elements. Note, however, that other functions could be used that give different weights to the individual off-diagonal elements.

In order to compare different mixing matrices it is useful to introduce some normalization. For example, we should consider the strength of a mixing matrix

$$\begin{pmatrix} 4 & 1 \\ 2 & 4 \end{pmatrix} \tag{3}$$

to be equal to

$$\begin{pmatrix} 1 & 0.25 \\ 0.5 & 1 \end{pmatrix} \tag{4}$$

as only the relative magnitude of the off-diagonal elements to the diagonal elements are important. Furthermore, because ICA algorithms can only estimate a mixing matrix up to a factor for each source signal, we are normalizing each column vector separately. Thus, we define the mixing strength to be

$$E_1^{\mathrm{M}} = \frac{1}{n(n-1)} \sum_{j=1}^{n} (\sum_{i=1}^{n} \frac{|m_{ij}|}{\max_k |m_{kj}|} - \frac{|m_{jj}|}{\max_k |m_{kj}|}), \tag{5}$$

where m_{ij} are the elements of the (typically unknown) mixing matrix $\mathbf{M}$. In addition to the normalization of each column of the matrix, we take also a normalization with respect to the rank n of the matrix into account so that the possible values of this strength measure range between 0 and 1 for mixing problems with arbitrary number of signals n.

2.3 The minimal mixing strength: $E_1^{\min}$

As mentioned above, ICA algorithms can estimate a linear mixing matrix only up to possible column permutation. However, the above definition of the mixing strength is sensitive to the permutations in the estimated mixing matrix. For example, consider the mixing matrix

$$\begin{pmatrix} 1 & 0.1 \\ 0.2 & 1 \end{pmatrix} \tag{6}$$

which has a mixing strength value of $E_1^{\mathrm{M}} = 0.15$, while the column-permutated matrix

$$\begin{pmatrix} 0.1 & 1 \\ 1 & 0.2 \end{pmatrix} \tag{7}$$

has a mixing strength value of $E_1^{\mathrm{M}} = 1$. The ICA estimates of the mixing matrix are thus not sufficient to estimate the strength of the mixing.

However, it is possible to calculate a lower bound of the possible mixing strength from the ICA estimates by exploring all possible permutations of $\mathbf{W}$. The main purpose of the estimation of the mixing strength in this paper is to explore the hypothesis that there is considerable mixing in real world datasets. Thus, the appropriate conservative measure for our hypothesis testing is

$$E_1^{\min} = \min\{E_1^{\mathrm{M}}(\mathbf{B})|\mathbf{B} \in P_c(\mathbf{A})\} \tag{8}$$

where $P_c(\mathbf{A})$ is a set of all matrices resulting from all possible column permutations of $\mathbf{A}$. However, calculating $E_1^{\min}$ in datasets with large number of features (signals) is not practical due to the large number ($n!$) of possible permutations. We therefore explore other possible estimates below.

2.4 Conflict case

There is an alternative approach to the above exhaustive search for the minimal mixing strength. The minimal mixing strength corresponds to the matrix $\mathbf{B} \in P_c(\mathbf{A})$ in which the sum of the diagonal elements is maximal. This matrix is easy to find in the case that each column vector has the maximum value at a position different from the position of the maximum values in all the other column vectors. This matrix can then be found by placing each column vector at the position of the index of the maximal element. This ordering can be done in quadratic time ($O(n^2)$).

However, initial experiments indicated that many estimated mixing matrices from the experiments described below did not have columns with unique positions of the maximal value. The case when two column vectors from one mixing matrix have maximal elements at the same index is termed a *conflict* in the following. for example, the matrix

$$\begin{pmatrix} 2 & 3 & 1 \\ 1 & 1 & 4 \\ 1 & 1 & 1 \end{pmatrix} \tag{9}$$

has a conflict number of $n_c = 2$, as two column vectors have the maximum on the at the first position. In this case we have still to test all permutations of the conflicting columns to find the minimal mixing strength. This operation is then proportional to $n_c!$. We report below on some experiments that show that the number of conflicts can be quite large (see Table 1). Thus, this methods is often not practical.

3 An Approximation of Minimal Mixing Strength

3.1 Lower bound on minimal mixing: E_1^{C}

We introduce here a new measure, called E_1^{C}, which is a lower bound on the minimal mixing strength,

$$E_1^{\mathrm{C}} \leq E_1^{\mathrm{min}}. \tag{10}$$

It is given by the sum of all elements of the matrix minus the largest element of each column,

$$E_1^{\mathrm{C}} = \frac{1}{n(n-1)} \sum_{j=1}^{n} (\sum_{i=1}^{n} \frac{|a_{ij}|}{\max_k |a_{kj}|} - 1). \tag{11}$$

The label 'C' of this measure indicates that the normalization and inner summation is carried out over the columns of the matrix. This measure does not dependent of either column or row permutations and can be calculated in quadratic time ($O(n^2)$).

It is easy to see that this measure is a lower bound of E_1^{min}. The measure E_1^{C} is equal to E_1^{min} if A has no conflicts because then the maximal element, which is 1 after normalization, can be placed on the diagonal with the ordering of the columns. If A has conflicts, then the above measure corresponds to the case of ignoring the conflicts and allowing each column vector being optimally placed with the maximal element on the diagonal. This introduces an error for each but one conflicting column, underestimating the true mixing strength because a value of one instead of a true diagonal element less than one is subtracted from the sum of all elements of the column vector in the measure E_1^{C}. In other words, compared to the measure $E_1^{\min}$, where the true diagonal element is subtracted, an error of $1 - a_{ii}$ is made for each but one conflicting column, where a_{ii} is the diagonal element of the permutated matrix with the smallest mixing strength.

3.2 A Better Measure: E_1 and E_1^{N}

A large value of E_1^C indicates a large value of $E_1^{\min}$ so that this measure is sufficient for our argument if E_1^C is large. However, a small value of E_1^C can still be caused by matrices with large $E_1^{\min}$ in the case of large number of conflicts. In this section we introduce a further measure which is a better estimate of the more likely mixing strength, although not strictly a lower bound of $E_1^{\min}$.

We first define the quantity E_1^{R},

$$E_1^{\mathrm{R}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\sum_{j=1}^{n} \frac{|a_{ij}|}{\max_k |a_{ik}|} - 1), \tag{12}$$

which is similar E_1^C except that the normalization is performed on the row vectors of the matrix $\mathbf{A}$. E_1^R is also independent of permutations, and in case of no conflicts it holds that $E_1^{\mathrm{R}} = E_1^{\mathrm{C}} = E_1^{\min}$. In the case of conflicts, E_1^{R} is an upper bound on $E_1^{\min}$,

$$E_1^{\mathrm{R}} \geq E_1^{\min}. \tag{13}$$

As E_1^{C} is a lower bound on $E_1^{\min}$, and E_1^{R} is an upper bound, it is appropriate to take the average

$$E_1 = \frac{1}{2}(E_1^{\mathrm{C}} + E_1^{\mathrm{R}}) \tag{14}$$

as approximation of $E_1^{\min}$. This quantity correspond to the measure E_1 introduced by Amari et al. [8] up to a normalization factor $\frac{1}{n(n-1)}$.

Finally, an even better estimate of the mixing strength can be achieved by replacing the term E_1^{R} in the above definition with an estimate that performs the row normalization after a column normalization. Formally, we define a column-normalized matrix

$$\tilde{\mathbf{A}} = (\frac{a_{ij}}{\max_k |a_{kj}|}) \tag{15}$$

and use this in a new measure

$$E_1^{\mathrm{CR}} \quad = \quad \frac{1}{n(n-1)} \sum_{i=1}^{n} (\sum_{j=1}^{n} \frac{|\tilde{a}_{ij}|}{\max_k |\tilde{a}_{ik}|} - 1). \tag{16}$$

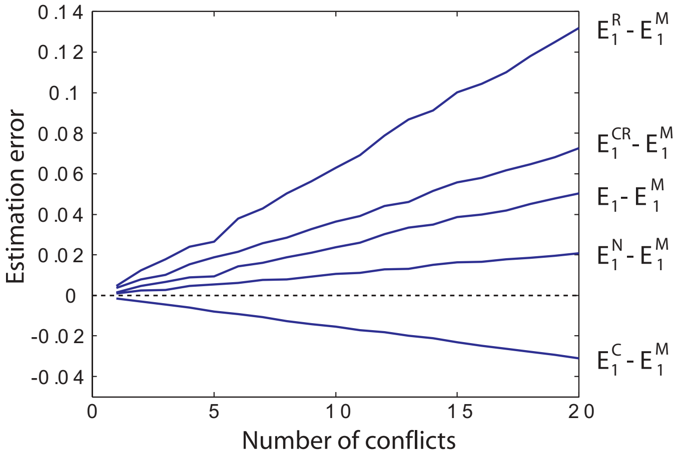

Figure 1: Dependency of various strength measures with the number of conflicts.

This matrix is still an overestimate of E_1^{M}, but the new estimate of the mixing strength

$$E_1^{\mathrm{N}} = \frac{1}{2}(E_1^{\mathrm{C}} + E_1^{\mathrm{CR}}), \tag{17}$$

is a better estimate than E_1 as demonstrated below. The superscript 'N' stands for 'normalized',

3.3 Example comparison of the strength measures

To demonstrate the different measure we performed a test with random mixing matrices. In this experiment a random mixing matrix of size 21 by 21 with elements drawn equally between 0 and 1 was added to a unit matrix. This matrix correspond to a mixing matrix without conflicts. To generate mixing matrices with conflicts we randomly picked a specific number of columns equal to the number of conflicts and exchanged the diagonal elements with the first element in the same column. The following experiments were performed on several such mixing matrices.

The results of the different measures relative to the mixing strength E_1^{M} is shown in Figure 1 as a function of the conflicts. The values represent averages over 30 runs with different mixing matrices. All the measures agree in case of no conflicts. However, the quality of the approximation of E_1^{M} is different for increasing number of conflicts. As analyzed above, E_1^{C} is always an underestimation of the true value in case of conflicts, and this difference is increasing linearly as the number of conflicts increases. E_1^{R} on the other hand is always an overestimate of E_1^{M}, and the difference to E_1^{M} is increasing with the number of conflicts. E_1 is slightly overestimating E_1^{M} but is a fairly good approximation of E_1^{M}. Finally, the best estimation of E_1^{M} in the presence of conflicts is the measure E_1^{N}.

We compared the various measures of the mixing strength here to E_1^{M}, which is possible as the mixing matrix is known in these experiments. In the context of estimating the mixing strength of real world datasets, we are, of course, more interested in the estimation of $E_1^{\min}$ as explained above. We did not calculate $E_1^{\min}$ in these experiments as this was prohibited by the computational complexity of this measure in the case of large numbers of conflicts. The minimal mixing strength, $E_1^{\min}$, is always smaller or equal to E_1^{M}. The measure of E_1^{N} is thus overestimating $E_1^{\min}$ compared to E_1^{M}. However, as mentioned above, E_1^{C} is always a lower bound on $E_1^{\min}$. Thus, E_1^{C} might overestimate the number of cases with small mixing strength, while E_1^{N} might underestimate the cases with small mixing strength. We are using therefore both measures, E_1^{C} and E_1^{N}, in the following study as the combination of these measures can provide a better picture of the possible range of expected mixing strength values.

4 Mixing in real world datasets

In the following we apply the measure E_1^{C} and E_1^{N} to estimate the mixing strength in real world datasets. Thirty-one datasets were chosen arbitrarily from four data collections, the StatLib-Datasets Archive [12], the Delve library [13], the UCI Machine Learning Repository [14], and the FMA collection [15]. We omitted datasets with missing data and non-numeric feature values for simplicity. In the remaining datasets we eliminated features which had no obvious problem-dependent meaning such as serial numbers, or which had obvious dependencies to other features such as classification numbers. Note that the used datasets came from a variety of subjects areas such as economics, robotics, or health informatics. The studied datasets have a number of features ranging from 2 to 33, and the number of samples in different datasets range from tens to thousands (see Table 1).

To estimate the mixing matrix (or a permutation thereof) we used standard ICA algorithms. Different ICA algorithms use different methods to estimate the statistical dependencies between signals to be minimized, and the different algorithms use also different minimization procedures. A ranking of those algorithms is difficult as each of them relies on specific assumptions and approximations that make them dependent on the specific signals under investigation. A comparison of several algorithms on mixtures of sub-gaussian, super-gaussian, and mixtures of sub- and super-gaussian signals has been done by Giannakopoulos et al. [16] (see also [17]). The main conclusion of these authors is that all algorithms can perform comparably well, although there are some differences in specific circumstances. A brief comparison by Li et al. [18] of three algorithms, the Infomax algorithm by Bell and Senjnowski [11] (augmented by Amari's natural gradient algorithm [8]), the fastICA algorithm by Hyrakunen et al. [7], and the JADE algorithm by Cardoso [10] found that FastICA and Jade performed similarly well, while the Infomax algorithm showed some difficulties in one of their experiments. Our own experiments showed a strong performance of the FastICA algorithm, which is computationally efficient, shows fast convergence,

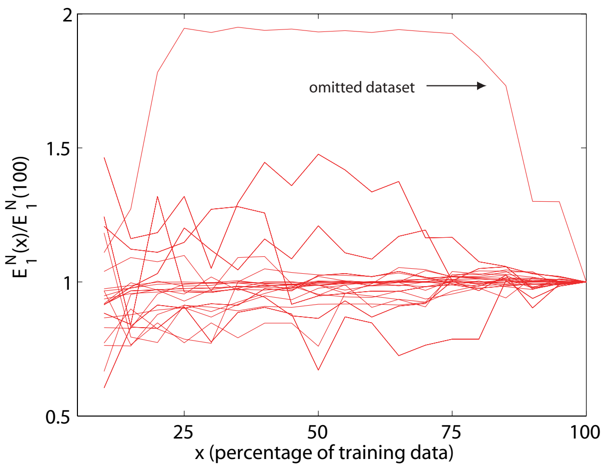

Figure 2: Dependency of the value E_1^{N} on the number of data used for the estimation in 31 datasets. Plotted is the value of E_1^{N} for different number of data points relative to the value as estimated from the complete dataset. The values represent averages over 20 ICA runs for each subset. A steady value for large coverage was taken as an indication for convergence of the estimation.

and is easy to use [19]. We used the deflation version of the FastICA algorithm with default parameters to calculate A in this study.

To verify the stability of the estimates with respect to the size of the samples we calculated the mixing strength E_1^{N} for different fractions of data. This is shown in Figure 2. Each curve represents the average of 30 trials. All but one dataset showed a stable estimate when most of the data was included, establishing some confidence that the number of samples is sufficient to estimate the mixing matrix. Only one dataset showed a strong variation of E_1^{N} for high percentages of the data. This dataset was not included in the following analysis.

The detailed results of the estimates of E_1^{C} and E_1^{N} are given in Table 1. We omitted attributes that were not specific problem-related, such as ID or entry numbers. The number of features are the remaining features used in the analysis. The results for the number of conflicts and the various mixing strength measures represent averages over 30 trials with different starting condition of the ICA algorithm. Datasets 1-17 are from the StatLib library [12], specifically (1) alr56, (2) alr57, (3) Boston house-price, (4) Body fat, (5) S&P Letters Data, (6) ch10, (7) ch17, (8) ch1a, (9) ch3a, (10) Wages, (11) Disclosure, (12) Irish Educational Transitions, (13) papir, (14) places, (15) pollen, (16) pollution, and (17) Child witness example. Datasets 18-25 are from the Delve library [13], specifically (18) KINematics-32fh, (19) KINematics-32fm, (20) KINematics-8fh, (21) KINematics-8fm, (22) PUMA DYNamics-32fh, (23) PUMA DYNamics-32fm, (24) PUMA DYNamics-8fh, (25) PUMA DYNamics-8fm. Datasets 26-28 are from the the UCI library [14], specifically (26) Liver-disorders Database,

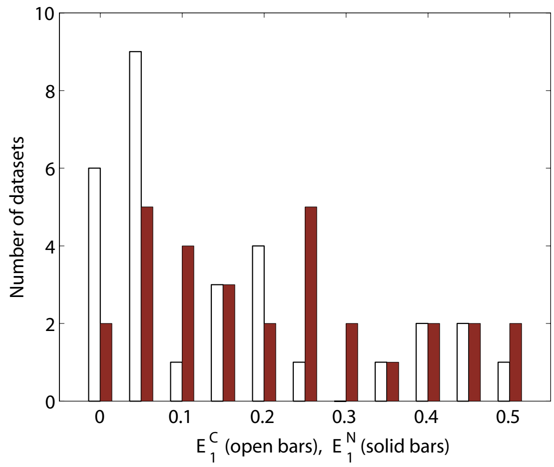

Figure 3: Distribution of E_1^C (open bars), which is a lower bound on the minimal possible mixing strength, and E_1^N (solid bars), which is a better estimate on the minimal possible mixing strength, in 30 real world datasets.

(27) Iris Plants Database, and (28) Wine recognition data. Datasets 29 and 30 are from the FMA library [15], specifically (29) Bank Data, and (30) Boston Stock.

A histogram of values E_1^{C}, which represent a strict lower bound on the minimal mixing strength E_1^{min}, is shown in Figure 3 (open bars). 24 out of the 30 datasets have a lower bound of the mixing strength larger than 0.025, while half of the dataset have values of E_1^{C} larger than 0.075. These results already indicate a large number of datasets have measurable mixing between the features, and it is possible that this number is even underestimated by the use of the measure E_1^C. We therefore compared the histogram of mixing strength estimations derived from E_1^C to the histogram of mixing strength estimation derived from E_1^{N} in Figure 3 (solid bars). With this estimate there are now 28 out of the 30 datasets with an estimated mixing strength larger than 0.025 and 23 out of 30 with an estimated mixing strength larger than 0.075. Interestingly, all of the 7 datasets with $E_1^{\mathrm{N}} < 0.075$ are from simulated robotrics experiments. The features in these datasets represent well designed measurements so that it can be expected that there is minimal redundancy (and hence mixing) in the feature values of these datasets.

5 Conclusions

The tests conducted in this study confirm that it is not uncommon that features in real world datasets have strong statistical dependencies. Hence, it is recommended to consider preprocessing of data with an independent component

Table 1: Results of the ICA analysis from 30 datasets from public machine learning datasets specified in the text. The measured quantities E_1^{N}, E_1^{C}, and the number of conflicts, are the average values over 30 runs with different initializations of the ICA algorithm.

SN	# Features	# Samples	# Conflicts	E_1^{N}	E_1^{C}
1	11	26	8.2	0.28	0.19
2	11	32	7.9	0.31	0.22
3	16	506	12.7	0.16	0.07
4	15	252	10.7	0.23	0.16
5	9	20640	7.9	0.16	0.03
6	7	60	1.1	0.44	0.44
7	13	68	11.9	0.11	0.04
8	4	704	3	0.24	0.05
9	4	50	0.9	0.5	0.49
10	11	534	7.1	0.21	0.16
11	4	662	3	0.16	0
12	6	500	4	0.28	0.21
13	13	29	4.3	0.42	0.41
14	9	329	5	0.27	0.21
15	5	481	2	0.42	0.39
16	16	60	14	0.11	0.03
17	14	42	5.4	0.44	0.43
18	33	8192	0	0.02	0.02
19	33	8192	0.3	0.02	0.02
20	9	8192	0.4	0.03	0.03
21	9	8192	1	0.1	0.09
22	33	8192	11.1	0.05	0.05
23	33	8192	13.8	0.06	0.05
24	9	8192	2	0.04	0.03
25	9	8192	1.6	0.06	0.04
26	6	345	1.4	0.2	0.17
27	4	126	2	0.52	0.42
28	13	138	11.4	0.11	0.02
29	3	60	2	0.29	0.03
30	2	35	1	0.37	0.2

analysis (ICA) for data mining, in particular when considering input variable selection [6].

Finding the minimal mixing strength in datasets with many features is complicated by the unknown order of possible source signals as ICA algorithms typically only produce estimates of mixing matrices up to permutation of columns of this matrix. Finding the mixing matrix with the smallest mixing strength from all possible permutations is computationally infeasible. We have introduced several measures that have a significantly reduced computational complexity which scales quadratically with the number of features. We showed that one of these measures, E_1^C, is a strict lower bound on the minimal mixing strength. We further introduced the measure E_1^N which is a good estimate of the expected mixing strength although not a lower bound on the minimal mixing strength.

The measure E_1^N is very similar to the measure E_1 introduced by Amari et al. [8]. However, note that the measure E_1 is most commonly used in the performance evaluation of ICA algorithms where the true mixing matrix is known. Here we adopted this measure to a situation where the true mixing matrix is unknown. Also, we augmented the original measure with a normalization factor to enable the comparison of mixing strength values between mixing matrices of different size. The measures E_1 and E_1^N are the same when the columns of the estimated mixing matrix are first normalized. This might be a common procedure in ICA studies, but our discussion show the importance of this step.

The paper demonstrates the usefulness of the ICA algorithm to estimate linear mixing matrices, and the new measures of a mixing strength estimation can be used to rapidly monitor some aspects of the nature of datasets that are considered for data mining purposes.

References

[1] L.A. Rendell and R. Seshu. Learning hard concepts through constructive induction: Framework and rationale. *Computational Intelligence*, 6(4):247–270, 1990.

[2] A.A. Freitas. Understanding the crucial role of attribute interaction in data mining. *Artificial Intelligence Review*, 16(3):177–199, November 2001.

[3] Sylvain Létourneau. *Identification of Attribute Interactions and Generation of Globally Relevant Continuous Features in Machine Learning.* PhD thesis, School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, Canada, August 2003.

[4] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[5] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, March 2003.

[6] Andrew D. Back and Thomas P. Trappenberg. Selecting inputs for modeling using normalized higher order statistics and independent component analysis. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 12(3):612–617, May 2001.

[7] Aapo Hyvärinen. Fast and robust fixed point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, April 1999.

[8] H.H.Yang S. Amari, A. Cichocki. A new learning algorihtm for blind source separation. *Advances in Neural Information Processing 8(Proc. NIPS'95)*, 10(3):626–634, April 1996.

[9] Aapo Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[10] Antoine Souloumiac Jean-Francois Cardoso. Blind beamforming for non gaussian signals. *IEE- Proceedings-F*, 140(6):362–370, December 1993.

[11] Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution source. *Neural Computation*, 7(6):1129–1159, November 1995.

[12] Statlib—datasets archive, http://lib.stat.cmu.edu/datasets/.

[13] Geoffrey Hinton Drew van Camp Michael Revow Zoubin Ghahramani Rafal Kustra Carl Edward Rasmussen, Radford M. Neal and Rob Tibshirani. Data for evaluating learning in valid experiments, http://www.cs.utoronto.ca/ delve/data/datasets.html.

[14] C.L. Blake and C.J. Merz. UCI repository of machine learning databases, http://www.ics.uci.edu/∼mlearn/mlrepository.html, 1998.

[15] Rob J Hyndman Spyros G. Makridakis, Steven C. Wheelwright. *Forecasting: Methods and Applications (3rd edition)*. John Wiley & Sons, 1998.

[16] Karhunen J. Giannakopoulos, X. and E. Oja. An experimental comparison of neural algorithms for independent component analysis and blind separation. *International Journal of Neural Systems*, 9(2):99–114, 1999.

[17] Xavier Giannakopoulos. Comparison of adaptive independent component analysis algorithms. Master's thesis, Helsinki University of Technology, Rakentajanaukio 2 C, March 1998.

[18] David Powers Yan Li and James Peach. Comparison of blind source separation algorithms. *Advances in Neural Networks and Appkications, World Scientific Engineering Society*, pages 18–23, 2000.

[19] Jie Ouyang. Improved icaivs algorithm with mutual information. Master's thesis, Dalhousie University, 2004.