SIGNAL DETECTION USING ICA: APPLICATION TO CHAT ROOM TOPIC SPOTTING

Thomas Kolenda, Lars Kai Hansen, and Jan Larsen

Informatics and Mathematical Modelling
Technical University of Denmark
Richard Petersens Plads, Bldg. 321
DK-2800 Kgs. Lyngby, Denmark
Emails: thko,lkh,jl@imm.dtu.dk

ABSTRACT

Signal detection and pattern recognition for online grouping huge amounts of data and retrospective analysis is becoming increasingly important as knowledge based standards, such as XML and advanced MPEG, gain popularity. Independent component analysis (ICA) can be used to both cluster and detect signals with weak a priori assumptions in multimedia contexts. ICA of real world data is typically performed without knowledge of the number of non-trivial independent components, hence, it is of interest to test hypotheses concerning the number of components or simply to test whether a given set of components is significant relative to a "white noise" null hypothesis. It was recently proposed to use the so-called Bayesian information criterion (BIC) approximation, for estimation of such probabilities of competing hypotheses. Here, we apply this approach to the understanding of chat. We show that ICA can detect meaningful context structures in a chat room log file.

1. INTRODUCTION

In [7] we initiated a development of a signal detection theory based on testing a signal for dynamic component contents. The approach is based on an approximate Bayesian framework for computing relative probabilities over a set of relevant hypotheses, hence obtaining control of both type I and type II errors. In this contribution we give additional detail and furthermore apply the approach to detection of dynamic components in a chat room log file.

Chats are self-organized narratives that develop with very few rules from the written interaction of a dynamic group of people and as a result often appear quite chaotic. When a new user enters a chat room a natural first action is to explore which topics that are being discussed.

Are chats simply a waste of time or is it possible to process chats to extract meaningful components? This paper is an attempt at modeling chat dynamics. There are numerous possible real-world applications of this type of analysis. One application is to provide retrospective segmentation of a chat log in order to judge whether a chat is worth engaging, another application would be to survey a large number of chats for interesting bits.

Our approach is unsupervised, with emphasis on keeping a low computational complexity in order to provide swift analyzes. This is achieved using a modified version of the *independent component analysis* (ICA) algorithm proposed by Molgedey and Schuster [14]. Our text representation is based on the vector space concept used e.g., in connection with *latent semantic analysis* (LSA) [2]. LSA is basically principal component analysis of text and LSA will be used for dimensional reduction in this work.

Independent component analysis of text was earlier studied in [8] and used for static text classification in [9]. We here extend this research to the ICA of text based on *dynamic* components [10].

Using the Molgedey and Schuster approach, the ICA solution is achieved by solving the eigenvalue problem of a quotient matrix, of the size of K^2, where K is the number of components. The notion of dynamic components was put forward by Attias and Schreiner [1]. Attias and Schreiner's approach is more general and includes both dynamic and non-linear separation, however, at the price of a considerably more complex algorithm and significantly longer estimation times than the approach used here. Comparisons between the two schemes in a neuroimaging context are provided in [16].

Molgedey and Schuster proposed an approach based on dynamic decorrelation which can be used if the independent source signals have different autocorrelation functions [14, 4, 6]. The main advantage of this approach is that the solution is simple and constructive, and can be implemented in a fashion that requires minimal user intervention (parameter tuning). In [6] we applied the Molgedey-Schuster algorithm to image mixtures and proposed a symmetrized version of the algorithm that relieves a problem of the original approach, namely that it occasionally produces complex mixing coefficients and source signals. In extension to this work we present a computational fast way of determining the lag parameter τ of the model.

2. PROBABILISTIC MODELING

Let a set of hypotheses about the structure of a signal be indexed by $m = 0, \cdots, M$, where $m = 0$ is a null-hypothesis, corresponding data being generated by a white noise source. Bayes optimal decision rule (under 0/1 loss function) leads to the optimal model,

$$m_{\mathrm{opt}} = \arg\max_m p(m|X). \tag{1}$$

The probability of a specific hypothesis given the observed data X is denoted by $P(m|X)$, using Bayes' relation this can be written as,

$$P(m|X) = \frac{P(X|m)P(m)}{\sum_m P(X|m)P(m)}, \tag{2}$$

This work is funded by the Danish Research Councils through the THOR Center for Neuroinformatics and the Center for Multimedia. LKH wishes to thank Terry Sejnowski for hosting a visit to the Salk Institute, July 2000, where parts of this work was initiated.

where $P(X|m)$ is the evidence and $P(m)$ is the prior probability which reflects our prior beliefs in the specific model in relation to the other models in the set, if no specific belief is relevant we will use a uniform distribution over the set $P(m) = 1/(M+1)$.

A model will typically be defined in terms of a set of parameters θ so that we have a so-called generative model density (likelihood) $P(X|\theta, m)$, this density is often given by the observation model. We then have the relation

$$P(X|m) = \int P(X, \theta|m)\, d\theta = \int P(X|\theta, m)P(\theta|m)\, d\theta. \quad (3)$$

The $P(\theta|m)$ distribution carries possible prior beliefs on the level of parameters, often we will assume so-called vague priors that have no or little influence on the above integral, except making it finite in the case X is empty (i.e., $P(\theta|m)$ is normalizable).

The integral in equation (3) is often too complicated to be evaluated analytically. Various approximation schemes have been suggested, here we will use the Bayesian Information Criterion (BIC) approximation [12]. This approximates the integral by a Gaussian in the vicinity of the parameters that maximize the integrant (the so-called maximum posterior parameters θ^*). With this approximation the integral becomes

$$P(X|m) \approx P(X|\theta^*, m)P(\theta^*, m)\left(\frac{2\pi}{N}\right)^{d/2}, \quad (4)$$

where d is the dimension of the parameter vector and N is the number of training samples. High-dimensional models (large d) are exponentially penalized, hence, can only be accepted if they provide highly likely descriptions of data.

3. VECTOR SPACE REPRESENTATION OF TEXT

In the vector space model Salton [17] introduced the idea that text documents could be represented by vectors, e.g. of word frequency histograms, and that similar documents would be close in Euclidean distance in the vector space. This approach is principled, fast, and language independent. Deerwester et al. [2], suggested to analyze sets of documents by principal component analysis of the associated vectors and dubbed the approach latent semantic analysis (LSA). The eigenvectors of the co-variance matrix correspond to "typical" histograms in the document sets. Scatterplots of the projections of documents onto the most significant principal components form a very useful explorative means of spotting topics in text databases, see e.g., Figure 1. This was extended to both supervised and unsupervised probabilistic descriptions as in, e.g., [5, 11, 15]

For the LSA approach the temporal ordering of documents is arbitrary. So in order to explore dynamical aspects of text we generalize the vector space representation as illustrated in the upper part of Figure 3. We consider a single contiguous text. A set of pseudo documents are formed by extracting fixed length windows (number of words L) from the contiguous text. We will let windows overlap by 50%. Each text window (pseudo document) is then processed as in standard vector space representation, using a list of terms. In particular we filter the text to remove terms in a list of stop words and non-standard characters, see e.g., [9]. From each such filtered document a term histogram based on P terms is generated and normalized to unit length. The total set of window documents form the $P \times N$ term/document matrix denoted X

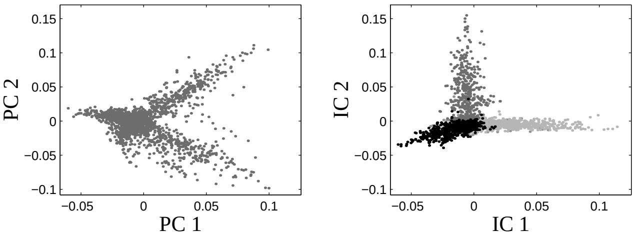


Fig. 1. The left panel shows a scatterplot of document in the LSA basis, while the right shows the similar plot based on the ICA components. The "ray-like" structure strongly indicates that we should use non-orthogonal basis vector in the decomposition of the data matrix. In the similar plot based on the ICA components the group structure is well-aligned with the axes. The gray shading of the dots show the ICA classification discussed later.

3.1. ICA Text Model

When applying ICA technology to text analysis we seek a decomposition of the term/document matrix,

$$X = AS, \qquad X_{j,t} = \sum_{k=1}^{K} A_{j,k} S_{k,t}, \quad (5)$$

where $X_{j,t}$ is the frequency of the j'th term in document (time) t, A is the $P \times K$ mixing matrix, and $S_{k,t}$ the k'th source signal in document (time) t.

The interpretation is that the columns of the mixing matrix A are "standard histograms" defining the weighting of the terms for a given context, while the independent source signals quantify how strongly each topic is expressed in the t'th document. Here we are specifically interested in temporally correlated source signals in order to detect contexts that are active in the chat room for extended periods of time.

4. MOLGEDEY SCHUSTER SEPARATION

Let $X_\tau = \{X_{j,t+\tau}\}$ be the time shifted data matrix. The delayed correlation approach for square mixing matrix is based on solving the simultaneous eigenvalue problem for the correlation matrices $X_\tau X^\top$ and $XX^\top$. Originally it was done in [14] by solving the eigenvalue problem of the quotient matrix $Q \equiv X_\tau X^\top (XX^\top)^{-1}$, but having a none square mixing matrix we need to extend the algorithm, see [6] for a more detailed derivation.

Using *singular value decomposition* (SVD) to find the principal component subspace, we decompose $X = UDV^\top$, where U is $P \times N$, D is $N \times N$, and V is $N \times N$ when assuming $P > N$ (in the case of text mining $P \gg N$). The quotient matrix can now be written as,

$$\begin{aligned} \widehat{Q} &\equiv \frac{1}{2} D(V_\tau^\top V + V^\top V_\tau) D^{-1} && (6) \\ &= \Phi \Lambda \Phi^{-1}. && (7) \end{aligned}$$

We have an option here for *regularization* by projecting onto small K-dimensional latent space by reducing the dimension of the SVD, i.e. also reducing the number of sources. U is then $P \times K$ and constituted by the first K columns, likewise, S is $K \times K$, and V

gets $N \times K$. In Φ we have the rectangular ($K \times K$) ICA basis projection onto the PCA subspace and U holds the projection from the PCA subspace to the original term/document space. The estimates of the mixing matrix and the source signals then are given by,

$$A = U\Phi, \tag{8}$$
$$S = \Phi^{-1} D V^{\top}. \tag{9}$$

The BIC approach for ICA signal detection consists in performing Molgedey-Schuster separation for a range of subspace dimensions, $K = 0, \cdots, K_{\max}$ and compute the approximate probabilities over the corresponding $K+1$ hypotheses in accordance with Eqs. (2),(4). In [7] simulation examples in fact showed that the BIC detector was more efficient than a test set based detector.

5. DYNAMIC COMPONENT LIKELIHOOD FUNCTION

PCA is used remove noise by projecting to a sparse latent space and thereby enhancing generalizability. We deploy the PCA model introduced in [3] and further elaborated in [13], where the signal space spanned by the first K eigenvectors has full covariance structure. The noise space $\mathcal{E}$ spanned by the remaining $P-K$ eigenvectors is assumed to be isotropic, i.e., diagonal covariance with noise variance estimated by $\sigma_\varepsilon^2 = (P-L)^{-1} \sum_{i=K+1}^{N} D_{ii}^2$. Assuming independence of signal and noise space we model

$$P(X|\theta, K) = P(Y|\Phi, K) P(\mathcal{E}|\sigma_\varepsilon^2). \tag{10}$$

where θ are model parameters, $Y = U^{\top} X$ is the signal space in which ICA is performed.

In order to compute the likelihood involved in the BIC model selection criterion Eq. (4) the likelihood for Y and $\mathcal{E}$ is required. It is easily verified [13] that

$$P(\mathcal{E}|\sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-N(P-K)/2} \cdot \exp(-N(P-L)/2) \tag{11}$$

The dynamic components S are assumed to be well described by their second statistics, hence, can be modeled by multi-variate normal colored signals, as would result from filtering independent, unit variance, white noise signals, by unknown, and source specific filters.

Given that no noise is present the following ICA model can be assumed where,

$$P(Y|\Phi, K) = \int dS \delta(Y - \Phi S) P(S). \tag{12}$$

The source distribution is given by,

$$P(S) = \prod_k \frac{1}{\sqrt{|2\pi\Sigma_s|}} \exp\left(-\frac{1}{2} \sum_{t,t'} S_{k,t} (\Sigma_s^{-1})_{t,t'} S_{k,t'} \right), \tag{13}$$

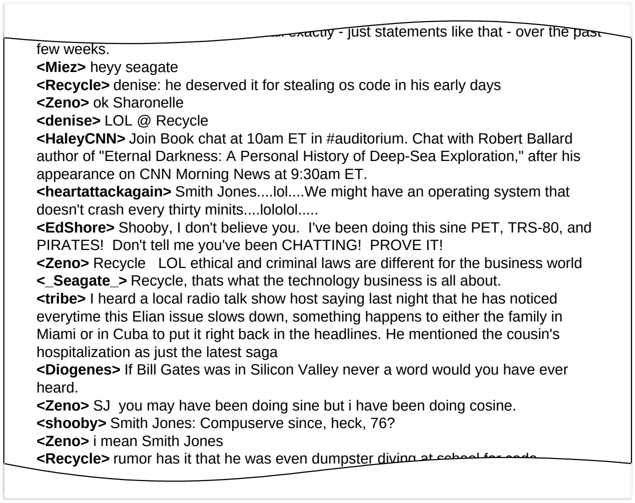
where the source covariance matrix is estimated as

$$\Sigma_{s_i} = \text{Toeplitz}([\gamma_{s_i}(0), ..., \gamma_{s_i}(N-1)]). \tag{14}$$

where $\gamma_{s_i}(m) = \sum_{n=1}^{N-m} s_i(n) s_i(n+m)$, $m = [0, \ldots, N-1]$, are estimated source autocorrelation functions, which form a Toeplitz matrices under the model assumptions. Evaluating the integral in equation (12) provides the expression

$$P(Y|\Phi, m) = \prod_k \frac{1}{\sqrt{|2\pi\Sigma_s|}} \left(\frac{1}{||\Phi||} \right)^N \exp\left(-\frac{1}{2} \sum_{t,t'} S_{k,t} (\Sigma_s^{-1})_{t,t'} S_{k,t'} \right). \tag{15}$$

with $||\Phi||$ being the absolute value of the determinant of Φ, while we use the notation $\widehat{S}_{k,t}$, for the sources estimated from A, Y $\widehat{S}_{k,t} = \sum_l (\Phi^{-1})_{k,l} Y_{l,t}$.



exactly - just statements like that - over the past few weeks.
<Miez> heyy seagate
<Recycle> denise: he deserved it for stealing os code in his early days
<Zeno> ok Sharonelle
<denise> LOL @ Recycle
<HaleyCNN> Join Book chat at 10am ET in #auditorium. Chat with Robert Ballard author of "Eternal Darkness: A Personal History of Deep-Sea Exploration," after his appearance on CNN Morning News at 9:30am ET.
<heartattackagain> Smith Jones....lol....We might have an operating system that doesn't crash every thirty minits....lololol.....
<EdShore> Shooby, I don't believe you. I've been doing this sine PET, TRS-80, and PIRATES! Don't tell me you've been CHATTING! PROVE IT!
<Zeno> Recycle LOL ethical and criminal laws are different for the business world
<_Seagate_> Recycle, thats what the technology business is all about.
<tribe> I heard a local radio talk show host saying last night that he has noticed everytime this Elian issue slows down, something happens to either the family in Miami or in Cuba to put it right back in the headlines. He mentioned the cousin's hospitalization as just the latest saga
<Diogenes> If Bill Gates was in Silicon Valley never a word would you have ever heard.
<Zeno> SJ you may have been doing sine but i have been doing cosine.
<shooby> Smith Jones: Compuserve since, heck, 76?
<Zeno> i mean Smith Jones
<Recycle> rumor has it that he was even dumpster diving at school for code

Fig. 2. The chat consists of a mixture of contributors discussing multiple concurrent subjects. The figure shows a small sample of the a CNN News Cafe chat line on April 5, 2000.

6. DETECTION OF DYNAMIC COMPONENTS IN CHAT

In this paper we present a retrospective analysis of a day-long chat in a CNN chat room. This chat is mainly concerned with a discussion of that particular days news stream. We show that persistent, and recurrent narratives in the form of independent dynamic components emerge during the day. These components have straightforward interpretations in terms of that particular days "top stories".

In conventional text mining the tasks of *topic detection and tracking* refer to automatic techniques for finding topically related material in streams of data such as newswire and broadcast news. The data is given as individual "meaningful units", whereas in the case of chat contributions are much less structured. The approach we propose for chat analysis share features with topic detection and tracking, however note that our approach is completely unsupervised.

The data set was generated from the daily chat at CNN.com in channel #CNN. In this particular chat room daily news topics are discussed by lay-persons. A CNN moderator supervises the chat to prevent non-acceptable contributions and to offer occasional comments.

All chat was logged in a period of 8.5 hours on April 5, 2000, generating a data set of 4900 lines with a total of 128 unique names participating. We do not know whether users logged on at different times with different names. The data set was cleaned by removal of non-user generated text, all users names, stop words and non-alphabetic characters. After cleaning the vocabulary consisted of $P = 2498$ unique terms.

The remaining text was merged into one string and a window of size 300 characters was used to segment the contiguous text in pseudo-documents. The window was moved forward approximately 150 characters between each line (without breaking words apart), leaving an overlap of approximately 50% between each window. A term histogram was generated from each pseudo-document forming the term/document matrix. We have performed a number of similar experiments with different window sizes, and also letting a document be defined simply as one user contribution. The latter provides a larger inhomogeneity in the document length. We found the best reproducibility for 300 character windows. This procedure produced a total of $N = 1114$ (pseudo-)documents.

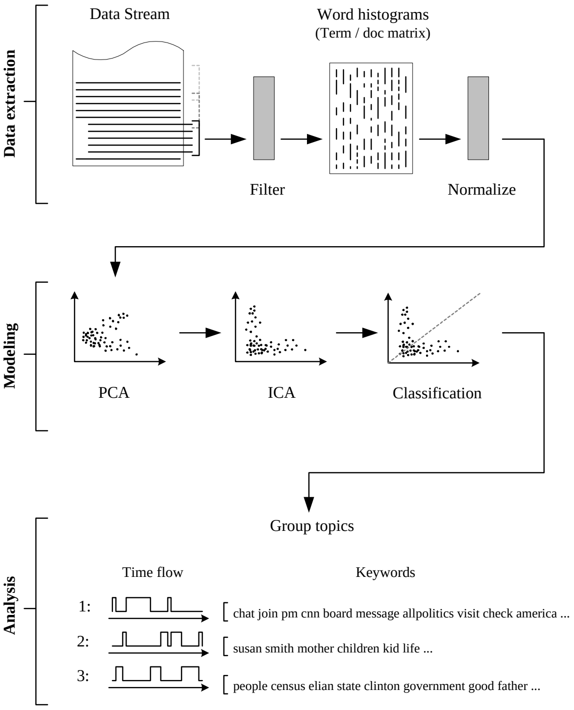

Data extraction
Data Stream
Word histograms
(Term / doc matrix)
Filter
Normalize
Modeling
PCA
ICA
Classification
Analysis
Group topics
Time flow
Keywords
1:
chat join pm cnn board message allpolitics visit check america ...
2:
susan smith mother children kid life ...
3:
people census elian state clinton government good father ...

Fig. 3. The text analysis process is roughly divided into tree phases: Data extraction and construction of the term histograms; modeling where the vector dimension is reduced and topics segregated and the analysis where the group structure is visualized and the dynamic components presented.

6.1. Optimal number of components

Using equation (1) we performed an exhaustive search for the optimal combination of the two parameters K, τ, leading to the optimal values $\tau = 169$, and $K = 4$. In figure 4 we show the spectrum of probabilities (2) over the set of hypotheses $K = 0 : 7$. We hereby detect a signal with four independent components.

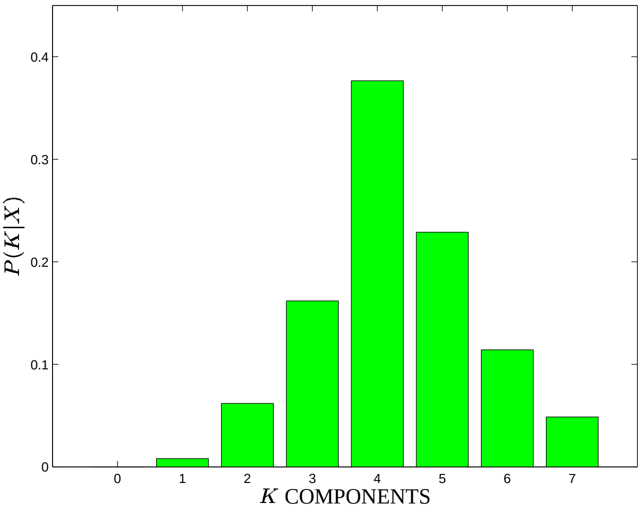


Fig. 4. Molgedey Schuster analyzes for $K = 0 - 7$ component with $\tau = 169$, for the chat data set. The most likely hypothesis contains four dynamic components.

6.2. Determination of τ

In numerous experiments with data of different nature it turned out that selection of the algorithm lag parameter τ is significant. A direct approach is to use equation (2) and test for all reasonable values of τ. This, however, does require a fairly large amount of computation and therefore not really attractive for online purposes.

As stated earlier, the ability of the algorithm to separate the dynamic components, is driven by exploiting the difference between the autocorrelations of the sources, $\gamma_{s_i}(m)$. Comparing the autocorrelations with the Bayes optimal model selection from (2), we observed a clear reduction in probability when the autocorrelation of the sources where overlapping, see Figure 5. Investigating this further, we formulated a objective function δ for identification of τ enforcing sources with autocorrelation values which are as widely distributed as possible.

For a specific τ, δ is given by,

$$\delta(\tau) = \sum_{i=1}^{K-1} |\rho_{s_{i+1}}(\tau) - \rho_{s_i}(\tau) - \frac{1}{K-1}|, \tag{16}$$

where $\rho_{s_{i+1}}(\tau) > \rho_{s_i}(\tau)$ are the sorted normalized autocorrelations $\rho_{s_i}(m) = \gamma_{s_i}(m)/\gamma_{s_i}(0)$.

Comparing the selection according to $\delta(\tau)$ and the Bayes optimal model selection procedure clearly showed identical behavior, as demonstrated in Figure 5.

The procedure for determination of τ thus consists of 1) estimating the sources and associated normalized autocorrelation functions for a initial value, e.g. $\tau = 1$. 2) Select the τ with the smallest $\delta(\tau)$, and reestimate the ICA. In principle this procedure is iterated until the value of τ stabilizes, which in experiments was obtained in less than 5 iterations.

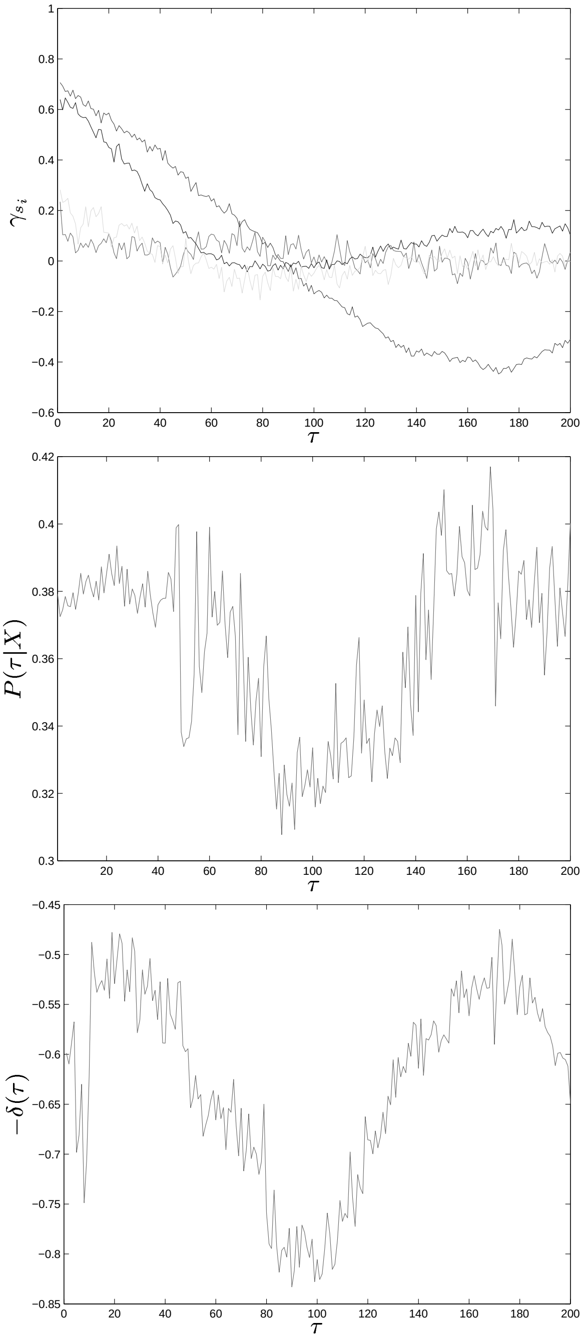


Fig. 5. The Bayesian scheme (middle) for estimating the optimal lag value τ is compared with a computationally much simpler approach (bottom), where the τ is chosen to be equal to the lag of which provides the most widely distributed autocorrelation function values of the sources (top). The best τ for was for the Bayesian approach was $\tau = 169$, and for the δ-function $\tau = 172$.

6.3. Retrospective event detection

In order to better understand the nature of the dynamic components we group the document stream. This is done by assuming equal

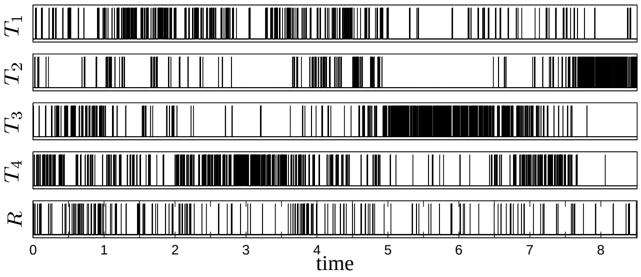


Fig. 6. The figure shows the result of the ICA classification into topic groups, as function of linear time during the 8.5 hours of chat. We used a simple magnitude based assignment after having normalized all components to unit variance, forming four topics $T_1, \cdots, T_4$ and a reject group R. The reject group was assigned whenever there was a small difference between the largest component and the runner up.

variance between components, and labeling a specific document s_i with $i = [1, \cdots, K]$ to the IC component closest in angle. In practice this amounts to selecting the index i as label for the component with largest magnitude, as shown in Figure 6, see [9] for further details. Note that the component sequences show both short and long time scale rhythms. To interpret the individual IC components found, we analyzed their normalized basis $(U\Phi)_i$ with $i = [1, \cdots, K]$ for the most 30% dominant words, that hereby made up a specific topic. The content of the topics spotted by this four-component ICA are characterized by keywords in Table 1. The first topic is dominated by the CNN moderator and immediate responses to these contributions, the second is a discussion on gun control, the third is concerned with the Susan Smith killings and her mother who appeared live on CNN, and the fourth is an intense discussion of the Cuban boy Elian's case. Hence, the approach has located topics of significant public interest.

	keywords
Topic 1	chat join pm cnn board message allpolitics visit check america
Topic 2	gun show
Topic 3	susan smith mother children kid life
Topic 4	people census elian state clinton government thing year good father time

Table 1. Projection of the independent components found in Figure 6 back to the term (histogram) space produces a histogram for each IC component. We select the terms with the highest back projections to produce a few keywords for each component. The first topic is dominated by the CNN moderator and immediate responses to these contributions. The second is a discussion on gun control. The third is concerned with the Susan Smith killings and her mother who appeared live on CNN, and finally the fourth is an intense discussion of the Cuban boy Elian's case.

7. CONCLUSION

By formulating independent component analysis in a Bayesian signal detection framework we can detect signals in complex multimedia signals with weak priors. Basically, we are detecting correlated structures against a white noise background. When applying the technology for analysis of a CNN chat log file we detected four interesting and highly relevant dynamic components that suggest that the approach may be of great help in navigating the web.

8. REFERENCES

[1] H. Attias and C.E. Schreiner: "Blind Source Separation and Deconvolution: The Dynamic Component Analysis Algorithm," *Neural Computation*, vol. 10, 1998, pp. 1373–1424.

[2] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K Landauer, and R. Harshman: "Indexing by Latent Semantic Analysis," *J. Amer. Soc. for Inf. Science*, vol. 41, 1990, pp. 391–407.

[3] L.K. Hansen and J. Larsen: "Unsupervised Learning and Generalization," in *Proceedings of IEEE International Conference on Neural Networks*, Washington DC, vol. 1, June 1996, pp. 25-30.

[4] L.K. Hansen and J. Larsen: "Source Separation in Short Image Sequences using Delayed Correlation," in P. Dalsgaard and S.H. Jensen (eds.) *Proceedings of the IEEE Nordic Signal Processing Symposium*, Vigsø, Denmark, 1998, pp. 253–256.

[5] L.K. Hansen, T. Kolenda, S. Sigurdsson, F. Nielsen, U. Kjems, and J. Larsen: "Modeling Text with Generalizable Gaussian Mixtures," *Proceedings of ICASSP'2000*, Istanbul, Turkey, vol. VI, 2000, pp. 3494–3497.

[6] L.K. Hansen, J. Larsen, and T. Kolenda: "On Independent Component Analysis for Multimedia Signals," in L. Guan, S.Y. Kung and J. Larsen (eds.) *Multimedia Image and Video Processing*, CRC Press, Chapter 7, 2000, pp. 175–200.

[7] L.K. Hansen, J. Larsen, and T. Kolenda: "Blind Detection of Independent Dynamic Components," *Proceedings of ICASSP'2001*, Salt Lake City, Utah, USA, SAM-P8.10, vol. 5, 2001.

[8] C.L. Isbell and P. Viola: "Restructuring Sparse High Dimensional Data for Effective Retrieval,"*Proceedings of NIPS98*, vol. 11, 1998, pp. 480–486.

[9] T. Kolenda, L.K. Hansen, and S. Sigurdsson: "Independent Components in Text," in M. Girolami (ed.) *Advances in Independent Component Analysis* Springer-Verlag, Berlin, 2000, pp. 229–250.

[10] T. Kolenda and L.K. Hansen: "Dynamical components of Chat," *Technical Report IMM, DTU*, ISSN 0909 6264 18-2000, 2000.

[11] J. Larsen, A. Szymkowiak, and L.K. Hansen: "Probabilistic Hierarchical Clustering with Labeled and Unlabeled Data," invited submission for *Int. Journal of Knowledge Based Intelligent Engineering Systems*, 2001.

[12] D.J.C. MacKay: "Bayesian Model Comparison and Backprop Nets," *Proceedings of Neural Information Processing Systems 4*, 1992, pp. 839–846.

[13] T.P. Minka: "Automatic Choice of Dimensionality for PCA," in *Porceedings of NIPS2000*, vol. 13, 2001.

[14] L. Molgedey and H. Schuster: "Separation of Independent Signals using Time-delayed Correlations," *Physical Review Letters*, vol. 72, no. 23, 1994, pp. 3634–3637.

[15] K. Nigam, A.K. McCallum, S. Thrun, and T. Mitchell: "Text Classification from Labeled and Unlabeled Documents using EM," *Machine Learning*, vol. 39, 2000, pp. 103–134.

[16] K.S. Petersen, L.K. Hansen, T. Kolenda, E. Rostrup, and S. Strother: "On the Independent Components in Functional Neuroimages," *Proceedings of ICA-2000*, Finland, June 2000.

[17] G. Salton: *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, 1989.