Cogito Componentiter Ergo Sum

Lars Kai Hansen and Ling Feng

Informatics and Mathematical Modelling,
Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
{lkh, lf}@imm.dtu.dk
www.imm.dtu.dk

Abstract. Cognitive component analysis (COCA) is defined as the process of unsupervised grouping of data such that the ensuing group structure is well-aligned with that resulting from human cognitive activity. We present evidence that independent component analysis of abstract data such as text, social interactions, music, and speech leads to low level cognitive components.

1 Introduction

During evolution human and animal visual, auditory, and other primary sensory systems have adapted to a broad ecological ensemble of natural stimuli. This long-time on-going adaption process has resulted in representations in human and animal perceptual systems which closely resemble the information theoretically optimal representations obtained by independent component analysis (ICA), see e.g., [1] on visual contrast representation, [2] on visual features involved in color and stereo processing, and [3] on representations of sound features. For a general discussion consult also the textbook [4]. The human perceptional system can model complex multi-agent scenery. Human cognition uses a broad spectrum of cues for analyzing perceptual input and separate individual signal producing agents, such as speakers, gestures, affections etc. Humans seem to be able to readily adapt strategies from one perceptual domain to another and furthermore to apply these information processing strategies, such as, object grouping, to both more abstract and more complex environments, than have been present during evolution. Given our present, and rather detailed, understanding of the ICA-like representations in primary sensory systems, it seems natural to pose the question: *Are such information optimal representations rooted in independence also relevant for modeling higher cognitive functions?* We are currently pursuing a research programme, trying to understand the limitations of the ecological hypothesis for higher level cognitive processes, such as grouping abstract objects, navigating social networks, understanding multi-speaker environments, and understanding the representational differences between self and environment.

Wagensberg has pointed to the importance of independence for successful 'life forms' [5]

> A living individual is part of the world with some identity that tends to become independent of the uncertainty of the rest of the world

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 446–453, 2006.

Thus natural selection favors innovations that increase independence of the agent in the face of environmental uncertainty, while maximizing the gain from the predictable aspects of the niche. This view represents a precision of the classical Darwinian formulation that natural selection simply favors adaptation to given conditions. Wagensberg points out that recent biological innovations, such as nervous systems and brains are means to decrease the sensitivity to un-predictable fluctuations. An important aspect of environmental analysis is to be able to recognize event induced by the self and other agents. Wagensberg also points out that by creating alliances agents can give up independence for the benefit of a group, which in turns may increase independence for the group as an entity. Both in its simple one-agent form and in the more tentative analysis of the group model, Wagensberg's theory emphasizes the crucial importance of *statistical independence* for evolution of perception, semantics and indeed cognition. While cognition may be hard to quantify, its direct consequence, human behavior, has a rich phenomenology which is becoming increasingly accessible to modeling. The digitalization of everyday life as reflected, say, in telecommunication, commerce, and media usage allows quantification and modeling of human patterns of activity, often at the level of individuals. Grouping of events or objects in categories is fundamental to human cognition. In machine learning, classification is a rather well-understood task when based on *labelled* examples [6]. In this case classification belongs to the class of *supervised* learning problems. Clustering is a closely related *unsupervised* learning problem, in which we use general statistical rules to group objects, without a priori providing a set of labelled examples. It is a fascinating finding in many real world data sets that the label structure discovered by unsupervised learning closely coincides with labels obtained by letting a human or a group of humans perform classification, labels derived from human cognition. *We thus define cognitive component analysis (COCA) as unsupervised grouping of data such that the ensuing group structure is well-aligned with that resulting from human cognitive activity* [7]. This presentation is based on our earlier results using ICA for abstract data such as text, dynamic text (chat), web pages including text and images, see e.g., [8,9,10,11,12].

2 Where Have We Found Cognitive Components?

Text Analysis. Symbol manipulation as in text is a hallmark of human cognition. Salton proposed the so-called vector space representation for statistical modeling of text data, for a review see [13]. A term set is chosen and a document is represented by the vector of term frequencies. A document database then forms a so-called term-document matrix. The vector space representation can be used for classification and retrieval by noting that similar documents are somehow expected to be 'close' in the vector space. A metric can be based on the simple Euclidean distance if document vectors are properly normalized, otherwise angular distance may be useful. This approach is principled, fast, and language independent. Deerwester and co-workers developed the concept of latent semantics based on principal component analysis of the term-document

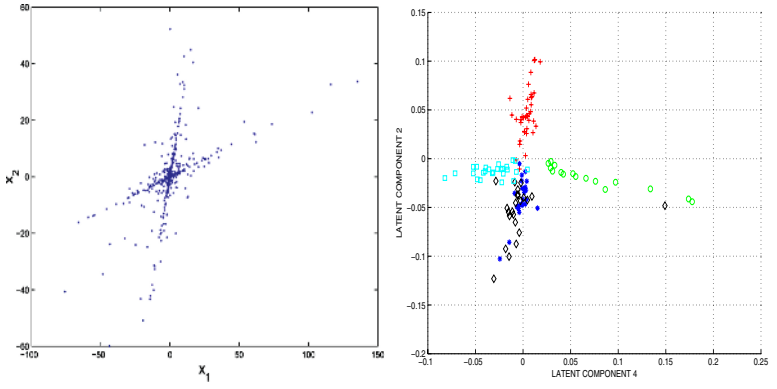

Fig. 1. Generic feature distribution produced by a linear mixture of sparse sources (left) and a typical 'latent semantic analysis' scatter plot of principal component projections of a text database (right). The characteristics of a sparse signal is that it consists of relatively few large magnitude samples on a background of small signals. Latent semantic analysis of the so-called MED text database reveals that the semantic components are indeed very sparse and does follow the laten directions (principal components). Topics are indicated by the different markers. In [16] an ICA analysis of this data set post-processed with simple heuristic classifier showed that manually defined topics were very well aligned with the independent components. Hence, constituting an example of cognitive component analysis: Unsupervised learning leads to a label structure corresponding to that of human cognitive activity.

matrix [14]. The fundamental observation behind the latent semantic indexing (LSI) approach is that similar documents are using similar vocabularies, hence, the vectors of a given topic could appear as produced by a stochastic process with highly correlated term-entries. By projecting the term-frequency vectors on a relatively low dimensional subspace, say determined by the maximal amount of variance one would be able to filter out the inevitable 'noise'. Noise should here be thought of as individual document differences in term usage within a specific context. For well-defined topics, one could simply hope that a given context would have a stable core term set that would come out as a eigen 'direction' in the term vector space. The orthogonality constraint of co-variance matrix eigenvectors, however, often limits the interpretability of the LSI representation, and LSI is therefore more often used as a dimensional reduction tool. The representation can be post-processed to reveal cognitive components, e.g., by interactive visualization schemes [15]. In Figure 1 (right) we indicate the scatter plot of a small text database. The database consists of documents with overlapping vocabulary but five different (high level cognitive) labels. The 'ray'-structure signaling a sparse linear mixture is evident.

Social Networks. The ability to understand social networks is critical to humans. Is it possible that the simple unsupervised scheme for identification of independent components could play a role in this human capacity? To investigate this issue we have initiated an analysis of a well-known social network of

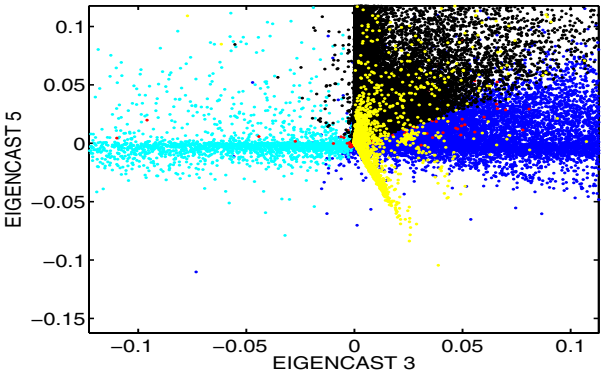

Fig. 2. The so-called actor network quantifies the collaborative pattern of 382.000 actors participating in almost 128.000 movies. For visualization we have projected the data onto principal components (LSI) of the actor-actor co-variance matrix. The eigenvectors of this matrix are called 'eigencasts' and they represent characteristic communities of actors that tend to co-appear in movies. The network is extremely sparse, so the most prominent variance components are related to near-disjunct sub-communities of actors with many common movies. However, a close up of the coupling between two latent semantic components (the region $\sim (0,0)$) reveals the ubiquitous signature of a sparse linear mixture: A pronounced 'ray' structure emanating from (0,0). The ICA components are color coded. We speculate that the cognitive machinery developed for handling of independent events can also be used to locate independent sub-communities, hence, navigate complex social networks.

some practical importance. The so-called *actor network* is a quantitative representation of the co-participation of actors in movies, for a discussion of this network, see e.g., [17]. The observation model for the network is not too different from that of text. Each movie is represented by the *cast*, i.e., the list of actors. We have converted the table of the about $T = 128.000$ movies with a total of $J = 382.000$ individual actors, to a sparse $J \times T$ matrix. For visualization we have projected the data onto principal components (LSI) of the actor-actor co-variance matrix. The eigenvectors of this matrix are called 'eigencasts' and represent characteristic communities of actors that tend to co-appear in movies. The sparsity and magnitude of the network means that the components are dominated by communities with very small intersections, however, a closer look at such scatter plots reveals detail suggesting that a simple linear mixture model indeed provides a reasonable representation of the (small) coupling between these relative trivial disjunct subsets, see Figure 2. Such insight may be used for computer assisted navigation of collaborative, peer-to-peer networks, for example in the context of search and retrieval.

Musical Genre. The growing market for digital music and intelligent music services creates an increasing interest in modeling of music data. It is now feasible to estimate consensus musical genre by *supervised* learning from rather short music segments, say 5-10 seconds, see e.g., [18], thus enabling computerized

handling of music request at a high cognitive complexity level. To understand the possibilities and limitations for unsupervised modeling of music data we here visualize a small music sample using the latent semantic analysis framework. The intended use is for a music search engine function, hence, we envision that a largely text based query has resulted in a few music entries, and the algorithm is going to find the group structure inherent in the retrieval for the user. We represent three tunes (with human genre labels: `heavy, jazz, classical`) by their spectral content in overlapping small time frames ($w = 30$msec, with an overlap of 10msec, see [18], for details). To make the visualization relatively independent of 'pitch', we use the so-called mel-cepstral representation (MFCC, $K = 13$ coefficients pr. frame). To reduce noise in the visualization we have further 'sparsified' the amplitudes. PCA provided unsupervised latent semantic dimensions and a scatter plot of the data on the subspace spanned by two such dimensions is shown in Figure 3. For interpretation we have coded the data points with signatures of the three genres involved. The ICA ray structure is striking, however, we note that the situation is not one-to-one as in the small text

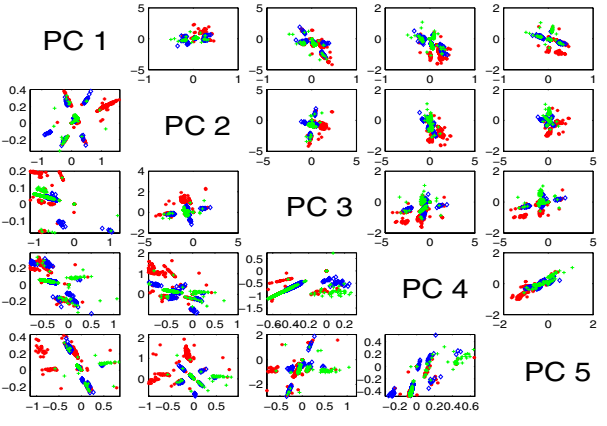

Fig. 3. We represent three music tunes (genre labels: `heavy metal, jazz, classical`) by their spectral content in overlapping small time frames ($w = 30$msec, with an overlap of 10msec, see [18], for details). To make the visualization relatively independent of 'pitch', we use the so-called mel-cepstral representation (MFCC, $K = 13$ coefficients pr. frame). To reduce noise in the visualization we have 'sparsified' the amplitudes. This was achieved simply by keeping coefficients that belonged to the upper 5% magnitude percentile. The total number of frames in the analysis was $F = 10^5$. Latent semantic analysis provided unsupervised subspaces with maximal variance for a given dimension. We show the scatter plots of the data of the first 1-5 latent dimensions. The scatter plots below the diagonal have been 'zoomed' to reveal more details of the ICA 'ray' structure. For interpretation we have coded the data points with signatures of the three genres involved: classical ($*$), heavy metal (diamond), jazz ($+$). The ICA ray structure is striking, however, note that the situation is not one-to-one (ray to genre) as in the small text databases. A component (ray) quantifies a characteristic musical 'theme' at the temporal level of a frame (30msec), i.e., an entity similar to the 'phoneme' in speech.

databases. A component quantifies a characteristic 'theme' at the temporal scale of a frame (30msec), it is an issue for further research whether genre *recognition* can be done from the salient themes, or we need to combine more than one theme to reach the classification performance obtained in [18].

Phonemes as Cognitive Components of Speech. There is a strong recent interest in representations and methods for computational auditory scene analysis, see e.g., Haykin and Chen's review on the cocktail party problem [19]. Low level cognitive components of speech encompass language specific features such as phonemes and speaker's voice prints. Such features can be considered 'presemantic' and would be recognized by human cognition without comprehension of the spoken message. We have recently investigated such low-level features and found generalizable features using ICA representations [20,21], here we give an example of such analysis based on four simple utterances s, o, f, a. We analysed 40 msec windows of length (95% overlap). The windows were represented by 16 Mel-cepstrum coefficients. After variance normalization the features were sparsified based on energy zeroing windows of normalized magnitudes with a statistical $z < 1.7$. This threshold process retains 55from original features. LSI/PCA was performed on the sparsified feature coefficients to get the most variant PCA components. The results in figure 4 seem to indicate that cognitive components corresponding to the phoneme /ae/ which opens the utterances s and f, can be identified using linear component analysis. For more details on such analysis see [20,21].

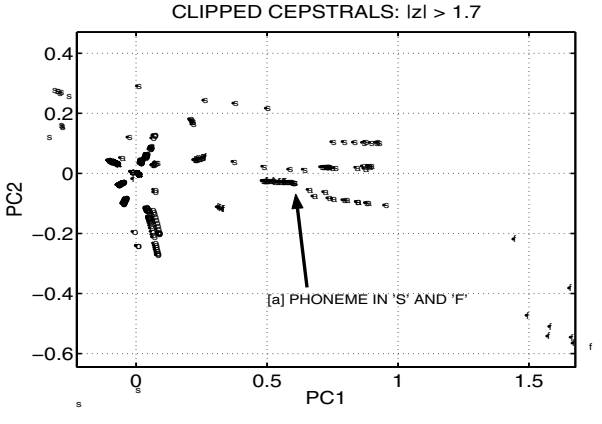

Fig. 4. Four simple utterances s, o, f, a were analysed. We analysed 40 msec windows of length (95% overlap). The windows were represented by 16 Mel-cepstrum coefficients. After variance normalization the features were sparsified based on energy zeroing windows of normalized magnitudes with a statistical $z \leq 1.7$. This threshold process retains 55% of the power in the original features. LSI/PCA was then performed on the sparsified feature coefficients for visualization. The results seem to indicate that generalizable cognitive components corresponding to the phoneme /ae/ opening the utterances s and f, can be identified using linear component analysis.

3 Conclusion

Cognitive component analysis (COCA) has been defined as the process of unsupervised grouping of data such that the ensuing group structure is well-aligned with that resulting from human cognitive activity. It is well-established that information theoretically optimal representations, similar to those found by ICA, are in use in several information processing tasks in human and animal perception. By visualization of data using latent semantic analysis-like plots, we have shown that independent components analysis is also relevant for representing semantic structure, in text and other abstract data such as social networks, musical features, and speech. We therefore speculate that the cognitive machinery developed for analyzing complex perceptual signals from multi-agent environments may also be used in higher brain function. Hence, we hypothesize that independent component analysis –given the right representation– may be a quite generic tool for COCA.

Acknowledgments

This work is supported by the Danish Technical Research Council, through the framework project 'Intelligent Sound', www.intelligentsound.org (STVF No. 26-04-0092). We thank our coworkers in the project for providing data for this presentation.

References

1. Bell, A.J., Sejnowski, T.J.: The 'independent components' of natural scenes are edge filters. Vision Research **37** (1997) 3327–3338
2. Hoyer, P., Hyvrinen, A.: Independent component analysis applied to feature extraction from colour and stereo images. Network: Comput. Neural Syst. **11** (2000) 191–210
3. Lewicki, M.: Efficient coding of natural sounds. Nature Neuroscience **5** (2002) 356–363
4. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
5. Wagensberg, J.: Complexity versus uncertainty: The question of staying alife. Biology and philosophy **15** (2000) 493–508
6. Bishop, C.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford (1995)
7. Hansen, L.K., Ahrendt, P., Larsen, J.: Towards cognitive component analysis. In: AKRR'05 -International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning, Pattern Recognition Society of Finland, Finnish Artificial Intelligence Society, Finnish Cognitive Linguistics Society (2005) Best paper award AKRR'05 in the category of Cognitive Models.
8. Hansen, L.K., Larsen, J., Kolenda, T.: On independent component analysis for multimedia signals. In: Multimedia Image and Video Processing. CRC Press (2000) 175–199

9. Hansen, L.K., Larsen, J., Kolenda, T.: Blind detection of independent dynamic components. In: IEEE International Conference on Acoustics, Speech, and Signal Processing 2001. Volume 5. (2001) 3197–3200
10. Kolenda, T., Hansen, L.K., Larsen, J.: Signal detection using ICA: Application to chat room topic spotting. In: Third International Conference on Independent Component Analysis and Blind Source Separation. (2001) 540–545
11. Kolenda, T., Hansen, L.K., Larsen, J., Winther, O.: Independent component analysis for understanding multimedia content. In et al. H.B., ed.: Proceedings of IEEE Workshop on Neural Networks for Signal Processing XII, Piscataway, New Jersey, IEEE Press (2002) 757–766 Martigny, Valais, Switzerland, Sept. 4-6, 2002.
12. Larsen, J., Hansen, L., Kolenda, T., Nielsen, F.: Independent component analysis in multimedia modeling. In ichi Amari et al. S., ed.: Fourth International Symposion on Independent Component Analysis and Blind Source Separation, Nara, Japan (2003) 687–696 Invited Paper.
13. Salton, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley (1989)
14. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. JASIS **41** (1990) 391–407
15. Landauer, T.K., Laham, D., Derr, M.: From paragraph to graph: latent semantic analysis for information visualization. Proc Natl Acad Sci **101** (2004) 5214–5219
16. Kolenda, T., Hansen, L.K., Sigurdsson, S.: Independent components in text. In: Advances in Independent Component Analysis. Springer-Verlag (2000) 229–250
17. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286** (1999) 509–512
18. Ahrendt, P., Meng, A., Larsen, J.: Decision Time Horizon For Music Genre Classification Using Short Time Features. In: EUSIPCO, Vienna, Austria (2004) 1293–1296
19. Haykin, S., Chen, Z.: The cocktail party problem. Neural Computation **17** (2005) 1875–1902
20. Feng, L., Hansen, L.K.: Phonemes as short time cognitive components. In: Submitted for ICASSP'06. (2005)
21. Feng, L., Hansen, L.K.: On low level cognitive components of speech. In: International Conference on Computational Intelligence for Modelling (CIMCA'05). (2005)