A Generative Score Space for Statistical Dialog Characterization in Social Signalling

Anna Pesarin¹, Marco Cristani^{1,2}, Paolo Calanca¹, and Vittorio Murino^{1,2}

¹ Dipartimento di Informatica, University of Verona, Italy

² Istituto Italiano di Tecnologia (IIT), Genova, Italy

Abstract. The analysis of human conversations under a social signalling perspective recently raised the joint attention of pattern recognition and psychology researchers. In particular, the dialog classification represents an appealing recent application whose aim is to go beyond the meaning of the spoken words, focusing instead on the way the sentences are pronounced by capturing natural (or hidden) characteristics, such the mood of the conversation. An effective strategy to face this issue is to encode the turn-taking dynamics in a generative model, whose structure is composed by conditional dependencies among first-order Markov processes. In this paper, we follow this strategy, investigating how to boost the classification performances of this model and of the related higher-order Markov extensions, through the definition of a novel generative score space. Generative score spaces are employed to increase generative classification in a discriminative way, also allowing a deep understanding of the processed data through the use of standard pattern recognition strategies. Experiments on real data certify the goodness of our intuition.

Keywords: social signalling, dialogue analysis, observed influence model.

1 Introduction

Social signal processing (SSP) aims at developing theories and algorithms that codify how human beings behave while involved in social interactions, putting together perspectives from sociology, psychology and computer science [1,2]. Here, the main entities to analyse are the social signals [2], i.e., temporal co-occurrences of social cues [3], that can be basically defined as a set of temporally sequenced changes in neuromuscular, neurocognitive and neurophysiological activity. Social cues are organized into five categories that are heterogeneous, multimodal aspects of a social interplay [2]: 1) physical appearance, 2) gesture and posture, 3) face and eyes behavior, 4) vocal behavior, and 5) space and environment.

The analysis of the social cues in the vocal behavior category is one of the issues most related to pattern recognition and machine learning themes. In general, this analysis consists in evaluating all the spoken cues that surround the verbal message and influence its actual meaning, characterizing, for example, particular social roles (*e.g.*, *dominance*, [4,5,6], *mirroring*, [7] and others [8]). A more recent challenge is to consider a conversation in its entirety, as a sample in a multidimensional space, in order to perform un/supervised clustering, indexing, retrieval and

other novel applications. For example, in [9], the goal was to predict the outcome of a specific conversational exchange finding and exploiting short meaningful portions of interaction. Another issue is that of the dialog classification (or characterization), aimed at capturing general aspects or characteristics of a dialog. For example, in [10], the main “atmosphere” (e.g., flat, aggressive, etc.) for a indefinitely long piece of dialogic exchange was classified, together with the capability of recognizing the presence in the dialog of a particular class of speakers, such as adults or children. It is worth noting that all these tasks usually do not involve speech recognition: actually, it happens very often that the meaning of the spoken sentences and the vocal behavior of a subject are completely in discordance.

In all the approaches above, generative models seem to be the main technique for exploiting vocal behavior cues for social signalling. Usually, they are employed for the analysis of the turn taking, *i.e.*, the sequence of turns in which a dialog participant can be in one of two states: silent or talking. Turn-taking dynamics may be effectively modelled as conditional dependencies among stochastic processes, where each process models the behavior of a single speaker. In particular, dynamic Bayesian networks were employed as efficient and expressive tools, especially, hidden Markov models (HMMs) and extensions [11], and influence model and extensions [12,7] (a.k.a. mixed memory Markov processes [13]). The common idea is to sample a dialog at fixed time intervals, to learn a representative model, and infer over the model parameters for detecting social aspects of that dialog. In [11], a two-layer HMM was employed to model individual and group action. In [12,7], the purpose was to detect the dominant interlocutor through social cues of mimicking. The authors employ an Observed Influence Model (OIM), *i.e.*, an aggregate of first-order Markov processes, each one addressing an interlocutor. OIM’s main feature is the capability of translating complex conditional dependencies among random variables with pairwise dependencies by means of weights called influence factors. Recently, in [10] a generative framework has been proposed, aimed at classifying a piece of conversation of variable length (from few minutes to hour), considering the nature of the people involved within (children, adults) and the main mood (flat, arguing). The framework is basically an OIM, fed by low-level auditory social signals, dubbed steady conversational periods (SCPs). They are built on duration of continuous slots of silence or speech, and, in addition, they take into account conversational turn-taking. In practice, SCPs allow to capture the attitude of self selecting for turn-taking even though the interlocutor has not yet completed his own turn. Further, they also indirectly model speech planning by characterizing the tendency to utter short sentences instead of longer propositions. We name here this generative framework as SCP model for brevity.

Employing generative machines for modelling dialog data is advantageous: the parameters of the model are intelligible, permitting inferences that highlight intuitively the nature of the data. For example, in [10], the coefficients of the transition matrices of the Markov model utilised suggest the dynamics of a process in a straightforward way, by identifying, for instance, highly probable or rare transitions.

In this paper, we focus on the dialog classification, considering the SCP model and proposing a principled way to boost its classification performances, also permitting, at the same time, a more informative analysis of the model.

To this end, we exploit a set of SCP models in an unconventional way, i.e., not as classifiers, but as feature ensembles: the idea is to build a novel generative score space [14], where the parameters of the SCP models are treated as features. More in detail, each sample (a dialog) of a given dialog class is used to learn an SCP model, and the model parameters can be considered as features in a joint multidimensional space. This embedding is repeated for all the classes of dialogs considered. Then, classical feature selection and ranking strategies are carried out, individuating the more discriminant features (i.e., parameters) for distinguishing the different dialog classes. Finally, discriminative classifiers are employed to perform the classification.

The concept of generative score space has recently raised the attention of the researchers, being a principled way to boost the classification performances of generative classifiers. In the majority of the approaches, discriminative techniques, like Support Vector Machines (SVMs), are fed with the (generative) features derived from the learned generative models, providing state-of-the-art performances [14,15,16,17]. The problem of such a formulation is that the discriminative part of the system hides one of the advantages brought from the generative modeling, i.e., we lose the intelligibility of the extracted features and, then, of the entire process.

In our framework, pre-processing feature selection and ranking strategies are carried out on the generative score space, allowing a full control and understanding of the collected features. In particular, we can observe what are the most useful features for classification, i.e., the most important parameters of the model that, in turn, means to highlight which transitions are more characteristic for a certain dialog class. In addition, we augment the complexity of the original generative framework, by embedding higher-order OIMs in the SCP model (i.e., considering Markov processes of higher order). This because another nice feature of the generative score space based approaches for classification is their ability to deal with overfitted models or with a small training sample size.

Summarizing, in this paper we reach three goals for dialog characterization: first, we perform classification in a very effective way; second, we employ more structured models for the dialog analysis, investigating whether higher order can encode a finer characterization of the turn taking; third, we understand the most important differences among different SCP models due to the embedding in the generative score space, realizing in a very intuitive way what are the behavioral patterns that characterize the different classes of dialogs.

The rest of the paper is organized as follows: in Sec. 2, mathematical recaps are given together with a brief description of the SCP model. Sec. 3 details our generative score space, and Sec. 4 reports the experimental results on a public dataset. Finally, Sec. 5 concludes the paper, summarising final observations on the turn-taking dynamics modelling and future perspectives of the work.

2 Mathematical Background

2.1 The Observed Influence Model

The observed influence model (OIM) is a simplified version of the influence model [12], that in practice operates on Markov processes instead of hidden Markov processes. Inheriting the notation of [10], the state variable of a Markov process is $S_t \in \{1, \dots, N\}$, and $P(S_t | S_{t-1}, \dots, S_{t-k})$ is the transition probability for a Markov model of order k . OIM factorizes the multi-process conditional relations among C Markov chains by means of a weighted linear combination of pairwise *inter-chain* and *intra-chain* transition probabilities. Considering first-order Markov chains with N states, the (full) factorization of the multi-process transition probability is

$$P({}^c S_t | {}^1 S_{t-1}, \dots, {}^C S_{t-1}) = \sum_{d=1}^C ({}^{c,d})\theta P({}^c S_t | {}^d S_{t-1}) \quad (1)$$

with $1 \leq c, d \leq C$, $({}^{c,d})\theta \geq 0$, $\sum_{d=1}^C ({}^{c,d})\theta = 1$. The value $P({}^c S_t | {}^d S_{t-1})$ represents the probability of going from state S_{t-1} of the chain d to state S_t of the chain c . The weight $({}^{c,d})\theta$ represents the influence that chain d exerts on chain c . A sketch of the model is depicted in Fig.1 a. A first-order influence model is thus defined as $\lambda = \{\{A^{(c,d)}\}, \Theta, \pi\}$, where $A^{(c,d)}$ is the *intra-chain* transition matrix when $c = d$, and represents the dynamics of a single process *per se*. When $c \neq d$, we consider the *inter-chain* matrices, modeling how much a state of a chain conditions the next state of the other chain. The $C \times C$ matrix Θ contains the influence weights, and π contains the (independent) initial probability distributions for all processes, *i.e.*, $\pi^{(c)} = \{\pi_i^{(c)}\}$ where $\pi_i^{(c)} = P(S_1^{(c)} = i)$.

The OIM transition factorization has space complexity $O(C^2 N^2 + C^2)$, where $C^2 N^2$ is due to the transition tables parameters, and C^2 to the influence coefficients. OIM learning of the $\{\theta\}$ coefficients is performed by standard constrained gradient descent [7,18], while the $\{\{A^{(c,d)}\}, \pi\}$ parameters are estimated by simple state counting.

2.2 The SCP Model

In their original framework [10], the authors focused on two-person conversations, where subjects 1 and 2 were captured in an appropriate environment, obtaining two synchronized separated audio sources. From the raw signals, a speech/silence thresholding was performed, obtaining a signal D , formed by two binary arrays $D^{(1)}$ and $D^{(2)}$, of length T .

Under this setting, a dialogue can be represented by an OIM, but the lack of synchronization between the start/end instants of the periods leads to problems in evaluating inter-chain conditional dependencies (see Fig.1 b).

Therefore, the authors propose the use of a turn taking-based feature, called *steady conversational period* (SCP), that is built on duration of continuous slots

Fig. 1. The generative framework: a) State factorization exploited in an Observed Influence Model. The area named Θ indicates the influence factors that apply to the state transitions, depicted as directed arrows. b) Synchronization through Steady Conversational Periods. We have two audio processes, $D^{(1)}$ and $D^{(2)}$, sampled at a given frequency, where audio samples are shown as *speech* (black dots) and *silence* (white dots) values. Continuous periods of speech or silence are not synchronized, so it is not possible to evaluate a first-order statistical transition probability among the periods. The global transitions (dashed red lines) define the SCPs so permitting to calculate first-order transition probabilities (black arrows).

of silence or speech. The extraction process of the SCPs assumes that whenever a process changes its state, it causes a *global* transition that affects also the opposite process, inserting a novel auto-transition state (see Fig.1 b). The fragmentation caused by global transitions synchronizes the processes, creating $\tilde{T} < T$ different SCPs ${}^cO_{\tilde{t}}$, where the apex $c \in 1, 2$ indexes the speaker and $\tilde{t} = 1, \dots, \tilde{T}$ enumerates the different SCPs.

The introduction of SCPs in the model makes it feasible to evaluate first-order intra- and inter-chain conditional probabilities (red dashed line in Fig.1 b). In order to take into account the different durations of each silence and speech segment, all SCPs related to the speech and to the silence are labelled into $\langle short, long \rangle$, after a Gaussian clustering over a training dataset.

More formally, given the clustering, each SCP ${}^cO_{\tilde{t}}$ takes one label among 1, 2, 3, 4, where 1, 2 address short and long continuous periods of speech, respectively, and the same applies with 3, 4 for the silence periods.

After that, an observed influence model $\lambda = \{\{A^{(c,d)}\}, \Theta, \pi\}$ is fitted to the SCP labels.

The parameters of the model intuitively indicate the conversational trend of each subject considered separately. The inter-chain transition parameters encode first-order state dependencies among processes, and influence factors mirror the influence that a process exerts on the other.

3 The Generative Score Space

In order to increase the classification accuracy of the generative framework, and, at the same time, to allow a discriminative analysis of the model parameters,

we build a generative score space $\mathcal{I}$. Following [19], such spaces can be built from the data available by considering each observed SCP sequence composed by the two synchronized SCP streams coming from the process 1 and 2 $O = (O_1, \dots, O_i, \dots, O_{\bar{T}})$, and a family of generative models $\mathcal{P} = \{P(O|\Psi_i)\}$ parameterized by Ψ_i .

The observed dialog O is mapped into a fixed-length score vector $\varphi_{\hat{F}}^f(O)$,

$$\varphi_{\hat{F}}^f(O) = \varphi_{\hat{F}} f(\{P_i(O|\Psi_i)\}), \quad (2)$$

where f is a function of the set of probability densities under the different models, and $\hat{F}$ is some operator applied to it. For instance, in case of the Fisher score [14], f is the log likelihood, and the operator $\hat{F}$ produces the first-order derivatives with respect to the parameters. Another example is the TOP kernel [15] for which the function f is the posterior log-odds and $\hat{F}$ is still the gradient operator.

In these cases, the generative score-space approaches help to distill the relationship between a model parameter θ_i and the particular data sample. After the mapping, a score-space metric must be defined in order to employ discriminative approaches.

In our case, f is the parameter extractor function (i.e., the function that estimates the parameters of a statistical distribution), and F is the identity operator. In practice, f extracts the transition parameters (by simple counting) and the influence coefficients (by gradient descent).

Given a set of M classes of dialogs, each formed by W sequences, the space $\mathcal{I}$, equipped with the traditional norm and Euclidean metric, could be seen thus formed by a set of multidimensional class-labeled samples; actually, on each sequence, a model is trained, that furnishes a set of features/parameters. Therefore, standard tools of data analysis can be applied. In our case, we want to highlight the discriminative power of the features in a classification context, and therefore we apply a feature selection (or ranking) strategy, and, subsequently, we apply different discriminative classifiers on the features subset. The feature selection/ranking strategies together with the discriminative classifiers employed will be detailed in the next section. Discriminative classifiers are preferred here, because they directly focus on estimating class posterior probabilities instead of modeling class distributions. Such classifiers should also be less affected by the curse of dimensionality problem.

In order to stress this aspect, and to assess how strong the improvement in the classification can be, when dealing with more structured models, we augment the order of the OIM embedded in the SCP-based generative framework. This is based on the following factorization:

$$P({}^cS_t|{}^1S_{t-1},\dots,{}^CS_{t-1},\dots,{}^1S_{t-k},\dots,{}^CS_{t-k}) = \sum_{d=1}^C ({}^{c,d})\theta P({}^cS_t|{}^dS_{t-1},\dots,{}^dS_{t-k}) \quad (3)$$

Encapsulating higher-order OIMs in the SCP-based generative framework is straightforward. The embedding in $\mathcal{I}$ leads to having the ensemble of features $\varphi_{\hat{F}}^f(x) = \{\{A^{(c,d)}\}, \Theta, \pi\}$, for each model (note that in this case, $A^{(c,d)}$ contains

N^{k+1} values). Considering in particular the number of parameters, we have $C^2 N^{k+1} + C^2$. For example, fixing $N = 4$ and $C = 2$ brings to 132 elements in the case of second order OIMs.

The rationale under the choice of this score space is that, employing parameters as discriminative features, we can understand what portions of a model differs from the other models at hand. For example, capturing the fact that a particular state transition is strongly discriminant for a certain class, means that such transition is peculiar for that model. This property cannot be mimicked by Fisher score based approaches, where the basic tool is the differentiation with respect to particular quantities (i.e., the log-likelihood in the Fisher score), that can suffer of the so-called “wrap-around” problem, where very different data points may map to the same derivative (see [17] for an example).

4 Experiments

In the experimental section, we employ the same database used for [10]¹, adopting the results reported in their paper as comparison. The code was written in MATLAB, and the classifier adopted, together with the feature selection strategies considered, were instantiated employing the PRTOOLS [20].

The corpus contains 41 dialogic conversations played by 30 subjects that can be grouped by age and mood in order to recognize three dialogue classes:

- C1: 13 flat semi-structured plus 5 flat unstructured dialogues between two adults ranging from 22 to 40 years.
- C2: 14 flat semi-structured dialogues between a child, ranging from 4 to 6 years, and an adult.
- C3: 9 arguing unstructured dialogues between two adults, ranging from 22 to 40 years

Each sample is approximately 10 min. long. In semi-structured conversation the moderator, a research-trained female psychologist who did not know the aim of the experiment, introduced in sequence 5 predetermined topics with fixed questions in a given order (school time, hobbies, friends, food, family). The class C3 was extracted by a corpus of phone office conversations driven by an operator who was aware of the experimental goal, and other subjects (Computer Science department employees) which were only warned about the possibility that an arguing issue might arise.

The classification task is performed into four different scenarios:

- (A) flat *vs* dispute - (*cat.1 vs cat.3*);
- (B) flat *vs* dispute, *general* - (*(cat.1 $\cup$ cat.2) vs cat.3*);
- (C) with *vs* without child - (*cat.2 vs cat.1*);
- (D) all *vs* all;

For the sake of clarity, let us suppose of having for each class L samples. In [10], the classification was performed in a generative way, using a Maximum Likelihood scenario, and cross-validating via leave-one-out (LOO). In other words, a

¹ The database is downloadable by contacting the authors.

class model was learned with L-1 samples, and testing was performed considering the last sample. In our case, we also employ LOO cross-validation, but we learn L-1 models for each class, one for each sample, projecting their scores into the generative score space, using the last sample as test. After that, we follow two different directions. First, we perform classification adopting all the features for each model, selecting different classifiers:

- *kernelc* [18]: a classifier based on a kernel or dissimilarity representation defined by Fisher approach;
- *knnc* [21]: a classifier based on k-nearest neighbor rule;
- *parzenc* [22]: a parzen classifier, using the best smoothing parameter of the kernel;

The best performances were reached by the *kernelc* classifier, and we report only these results in Fig.2 for brevity.

Case	1stOrder OIM (gen. class., [10])	1stOrder OIMGSS	
		Without Feat Sel	Forward Feat Sel
A	86%	89%	H=9 100%
B	86%	100%	H=7 100%
C	78%	93%	H=18 100%
D	80%	93%	H=27 97%

Case	2ndOrder OIM	2ndOrder OIMGSS	
		Without Feat Sel	Forward Feat Sel
A	93%	93%	H=5 100%
B	95%	97%	H=10 100%
C	76%	83%	H=10 92%
D	78%	73%	H=11 95%

Fig. 2. Classification results. In each table, the second column reports the results obtained by the generative approach, the third column shows the use of the *kernelc* classifier on all the features in the score space, the fourth column reports the results obtained after feature selection (H is the number of features considered).

Through the generative score space embedding, the classification performances augmented, except for the scenario *D* (all vs all) in the second order case. Investigating the feature space, we found that several parameters were shared among classes. Therefore, we employ forward feature selection (*ffs*) based on the 1-nearest-neighbor classification criterion. In this case, the capability of our score space to explain the data is evident. In all the cases, the generative performances were outperformed. Please note that each scenario required a different number of features for reaching the best performance. In particular, the scenario A and B were the simplest, and required a small set of features. Scenario C and D were more difficult, and a bigger number of feature were evaluated. It is worth noting how the second order model produced both a pure generative modeling and a score space that are less informative than those of order 1 (this considering the scenario C and D, which are more challenging wrt the first two). In this case, we can assume that the first order reasoning works better in this kind of scenario.

In order to understand the importance of all the selected features, we rank them, employing the *featrank* 1-nearest-neighbor feature ranking strategy, which evaluates the performances of each single feature taken separately. In this way, we highlight the parameters more discriminative for classification. For example, in the scenario A, in the first-order case, the transition probability between long speech state (SCP value = 4) of speaker 1 and long silence state of speaker 2 (SCP value = 2) was present as important feature (rank 1). This information serves to address a quantitative analysis about the nature of the models learned. In specific, the value of the above transition probability was high for the flat conversations (0.8), very low (0.1) for the arguing discussion. This mirrors the fact that a turn taking dynamics of a calm dialog implies that, especially after a long period of speech of a speaker, the other takes a while for thinking and elaborating its turn. In the arguing conversation, this dynamics is not present. Another important feature/parameter we found (rank 2) is that of a short speech of a subject (SCP value = 3) after a short speech of the other person. This transition has high probability for the arguing conversation, low for the flat, and witnesses the fact that in a fight, periods are usually shorter, and the speakers talk on each other.

5 Conclusions

In this paper, we propose a novel generative score space that operates directly on the parameters of a generative model, for increasing the classification performance on a social signal application. The peculiarity of our approach is to extract directly the parameters of the model, instead of relying on differentiating over the log-likelihood. This allows to highlight better the importance of the parameters, by means of feature selection/ranking strategies. This leads to higher classification performance thanks to discriminative reasoning on the selected features, and to understand better the data modeled. The future perspective is to better characterize theoretically such space, reasoning on expected classification bounds that can be achieved with it.

References

1. Pantic, M., Pentland, A., Nijholt, A.: Special issue on human computing. IEEE Trans. on Systems, Man, and Cybernetics, Part B 39(1) (2009)
2. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing 27(12), 1743–1759 (2009)
3. Ambady, N., Rosenthal, R.: Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological Bulletin 11(2), 256–274 (1992)
4. Jayagopi, D., Hung, H., Yeo, C., Gatica-Perez, D.: Modeling dominance in group conversations using nonverbal activity cues. Trans. Audio, Speech and Lang. Proc. 17(3), 501–513 (2009)
5. Choudhury, T., Basu, S.: Modeling conversational dynamics as a mixed-memory markov process. In: Saul, L.K., Weiss, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems 17, pp. 281–288. MIT Press, Cambridge (2005)

6. Vinciarelli, A.: Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Transactions on Multimedia* 9(6), 1215–1226 (2007)
7. Basu, S., Choudhury, T., Clarkson, B., Pentland, A.: Learning human interaction with the influence model. MIT MediaLab, Tech. Rep. 539 (2001)
8. Pentland, A.: Social signal processing. *IEEE Signal Processing Magazine* 24(4), 108–111 (2007)
9. Curhan, J., Pentland, A.: Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first five minutes. *Journal of Applied Psychology* 92, 802–811 (2007)
10. Cristani, M., Pesarin, A., Drioli, C., Perina, A., Tavano, A., Murino, V.: Auditory dialog analysis and understanding by generative modelling of interactional dynamics. In: Second IEEE Workshop on CVPR4HB, Miami, Florida (2009)
11. Zhang, D., Gatica-Perez, D., Bengio, S., McCowan, I., Lathoud, G.: Modeling individual and group actions in meetings with layered hmms. *IEEE Transactions on Multimedia* (May 2005)
12. Asavathiratham, C.: A tractable representation for the dynamics of networked markov chain. Ph.D. dissertation, Dept. of ECS, MIT (2000)
13. Saul, L., Jordan, M.: Mixed memory markov models: Decomposing complex stochastic processes as mixtures of simpler ones. *Machine Learning* 37(1), 75–87 (1999)
14. Jaakkola, T., Haussler, D.: Exploiting generative models in discriminative classifiers. In: Proceedings of the 1998 conference on Advances in neural information processing systems II, pp. 487–493. MIT Press, Cambridge (1999)
15. Tsuda, K., Kawanabe, M., Rätsch, G., Sonnenburg, S., Müller, K.: A new discriminative kernel from probabilistic models. *Neural Comput.* 14(10), 2397–2414 (2002)
16. Bicego, M., Pekalska, E., Tax, D., Duin, R.: Component-based discriminative classification for hidden markov models. *Pattern Recogn.* 42(11), 2637–2648 (2009)
17. Perina, A., Cristani, M., Castellani, U., Murino, V., Jojic, N.: Free energy score space. In: Advances in Neural Information Processing Systems, vol. 22, pp. 1428–1436 (2009)
18. Duda, R., Hart, P., Stork, D.: *Pattern Classification*. John Wiley and Sons, Chichester (2001)
19. Smith, N., Gales, M.: Speech recognition using svms. In: NIPS, pp. 1197–1204 (2001)
20. Duin, R., Juszczak, P., Paclík, P., Pekalska, E., DeRidder, D., Tax, D.: Prtools version 4.1: A matlab toolbox for pattern recognition. Internet (2004), <http://www.prtools.org>
21. Fukunaga, K.: *Statistical Pattern Recognition*, 2nd edn. Academic Press, London (1990)
22. Lissack, T., Fu, K.: Error estimation in pattern recognition via l-distance between posterior density functions. *IEEE Trans. Inform. Theory* 22, 34–35 (1976)