

APPLICATION OF t -CHERRY JUNCTION TREES IN PATTERN RECOGNITION

TAMÁS SZÁNTAI AND EDITH KOVÁCS

ABSTRACT. Pattern recognition aims to classify data (patterns) based either on a priori knowledge or on statistical information extracted from the data. In this paper we will concentrate on statistical pattern recognition using a new probabilistic approach which makes possible to select the so called 'informative' features. We develop a pattern recognition algorithm which is based on the conditional independence structure underlying the statistical data. Our method was successfully applied on a real problem of recognizing Parkinson's disease on the basis of voice disorders.

KEYWORDS: *pattern recognition, probabilistic modeling, h -uniform hypertree, t -cherry junction tree*

2000 Mathematics Subject Classification: 05C65, 62H30, 68T10, 62P10.

1. INTRODUCTION

In this paper we propose a probabilistic approach for statistical pattern recognition. The classifier will be achieved by supervised learning. The classification uses Bayes decision rule.

In our approach we estimate the underlying multi-variate probability distribution, by the exploitation of the conditional independencies between the variables (features). This is achieved by fitting to the training data a t -cherry junction tree (see [4], [8]). This represents the novelty of the approach introduced here.

A problem that occurs is that databases often contain redundant data, which mean a great number of features that may lead to overfitting of the model. This may cause poor generalization (the performance of the model on a new test data). We use the t -cherry junction tree approach to discover the Markov Random Field (MRF) underlying the variables. In an MRF each random variable depends on all others through the variables of its neighbourhood (see for example in [5]).

Throughout this paper we shall use the term of 'pattern' to denote a d -dimensional data vector $\mathbf{x} = (x^1, \dots, x^i, \dots, x^d)$ of measurements, where x^i is the value of the feature denoted by X^i . We assume that there exist M groups or classes that can be associated with each pattern. The categorical variable which takes values in $\{1, \dots, M\}$ is denoted by Y . The variables contained in the neighbourhood of Y are called informative variables for Y . Our method can be used for discovering those variables (features) which are informative for the categorical variable Y .

In the second section we describe how the t -cherry junction tree approach can be applied to the classification problem. In the third section we deal with a real world problem. We use our approach for selecting features of voice that are able to predict whether a patient has Parkinson's disease or not.

2. THE CLASSIFICATION METHOD BASED ON t -CHERRY JUNCTION TREE

A classifier is constructed on the basis of a training set $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ and is denoted by g_n . For a given $\mathbf{x} \in R^d$ the value $y \in \{1, \dots, M\}$ of Y is guessed by $g_n(\mathbf{x}; (\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n))$.

The construction of g_n in this way is called supervised learning.

We assume that $(\mathbf{X}_1, Y_1)^T, \dots, (\mathbf{X}_n, Y_n)^T$ is a sequence of independent identically distributed random vectors having the same distribution as $(\mathbf{X}, Y)^T$.

The training dataset may be the result of experimental observation (meteorological data, ECG data...) The y_i 's could be obtained through measurements or through expert who filled y_i 's after having x_i 's.

In order to find the best fitting k -width junction tree, we search in the class of k -width t -cherry junction trees. We intend to find the approximation which minimizes the Kullback-Leibler divergence of the approximation and the true probability distribution. The method for finding a k -th order t -cherry junction tree, using this criterion is presented in [8].

On the basis of the constructed t -cherry junction tree we select the informative features by the following algorithm.

Algorithm. Selection of the informative features.

1. Give as input the training data set in discretized form.
2. From the empirical probability distribution determine all the k -th order marginals (it is favorable if k is not to large (less than 7, say).
3. Find the heaviest weighted tree, in sense (see [8]):

$$\sum_{(X^{i_1}, \dots, X^{i_c}) \in \mathcal{C}} I(X^{i_1}, \dots, X^{i_c}) - \sum_{(X^{j_1}, \dots, X^{j_s}) \in \mathcal{S}} (\nu_{j_1, \dots, j_s} - 1) I(X^{j_1}, \dots, X^{j_s}) \rightarrow \max,$$

where $I(X^{i_1}, \dots, X^{i_c})$ is the information content (see [1]) of $P(X^{i_1}, \dots, X^{i_c})$.

4. Output : the set of clusters \mathcal{C} and the set of separators \mathcal{S} .
5. Select those clusters which contain the variable Y .
6. Select those variables which occur in the clusters selected in step 5 as informative variables.

For the set of informative features we introduce the notation $X_{\text{inf}} = \{X^{i_1}, \dots, X^{i_r}\}$. The joint probability distribution of the random vector $(\mathbf{X}_{\text{inf}}, Y)^T$ can be expressed using those marginals only, which correspond to the clusters that contain Y .

For the classifier g_n we give a formula which depends on the probability distribution of $(\mathbf{X}_{\text{inf}}, Y)^T$. This way for classifying a new d -dimensional vector we use the informative features only.

3. RECOGNIZING PARKINSON'S DISEASE FROM VOICE DISORDERS

Voice disorders can be premonitory of different diseases. This observation makes possible new ways of investigations and diagnosis. The idea to check up our method in discovering the connection between voice disorders and having

Parkinson's disease (PD) came from [6]: "Research has shown that approximately 90% of people with PD exhibit some form of vocal impairment [3], [7]. Vocal impairment may also be one of the earliest indicators for the onset of the illness [2]".

The dataset was created by Max Little of the University of Oxford, in collaboration with the National Center for Voice and Speech, Denver, Colorado, who recorded the speech signals, and provided by UCI Machine Learning Repository on the internet: <http://archive.ics.uci.edu/ml/datasets/Parkinsons>.

This dataset is composed of a range of biomedical voice measurements from 195 voice recording of 31 people, 23 with Parkinson's disease (PD). The main aim of the data is to discriminate healthy people from those with PD.

We choose randomly from the data a test set which contains 10 recordings of healthy people and 20 from ill ones. The rest of 165 vectors remained the training data set.

Only the cluster $(X^1, X^{10}, X^{17} = Y, X^{19}, X^{20})$ was informative. The informative features are: Fo, shimmer, DFA, spread1. Using the probability distribution $P(X^1, X^{10}, X^{19}, X^{20}, Y)$, we obtained a 93% correct classification performance for the classifier. The authors of paper [6] got 91.4 % correct classification performance using ten features and a Kernel support vector machine.

4. CONCLUSIONS

In this paper we approximate the multivariate probability distribution by exploiting the conditional independencies between the features. The construction is based on the fitting of a t -cherry junction tree to the training data set. The advantage of the method presented is that it makes possible the selection of informative features. Then we define a classifier which uses just the informative features. This way we can reduce the dimensionality of the problem and get a good generalization. The numerical results presented confirm the efficiency of our approach.

References

- [1] T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley Interscience, New York, (1991)

- [2] J. R. Duffy "Motor Speech Disorders: Substrates Differential Diagnosis, and Management , 2nd ed. St Louis, MO: Elsevier, 2005.
- [3] A. K. Ho, R. Iansek, C. Marigliani, J.L. Bradshaw, and S. Gates, Speech impairment in a large sample of patients with Parkinson's disease, *Behav. Neurol.*, 11, (1998), 131-137.
- [4] E. Kovács and T. Szántai, On the Approximation of Discrete Multivariate Probability Distribution using the new concept of t -cherry junction tree, *Lecture Notes in Economics and Mathematical Systems, 633, Proceedings of the IFIP/IIASA/GAMM Workshop on Copying with Uncertainty, Robust Solutions, 10-12 December, 2007, IIASA, Laxenburg*, Springer Verlag, Heidelberg, 2010, 39–56.
- [5] E. Kovács and T. Szántai, Some improvements of t -cherry Junction Trees, in: 2008 First International Conference on Complexity and Intelligence of the Artificial and Natural Complex Systems. Medical Applications of the Complex Systems. Biomedical Computing, CANS 2008, Marosvásárhely, Romania, 2008.11.08-2008.11.10, eds. Iantovics B., Enachescu C., Filip F. G., IEEE Computer Society, Los Alamitos, 2009, 117-129.
- [6] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig, Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease, *IEEE Transactions on Biomedical Engineering*, 56, No. 4, April, (2009), 1015-1022.
- [7] J.A. Logemann, H.B. Fisher, B. Boshes, E. R. Blonsky, Frequency and co-occurrence of vocal-tract disfunctions in speech of a large sample of patients with Parkinson's disease, *J. Speech, Hear. Disord.*, 43, (1978) 47-57.
- [8] T. Szántai and E. Kovács, Hypergraphs as Means of Discovering the Dependence Structure of a Discrete Multivariate Probability Distribution, *Proc. Conference APplied mathematical programming and Modelling (APMOD), 2008, Bratislava, 27-31 May, 2008, Annals of Operations Research*, to appear.

Tamás Szántai
Institute of Mathematics
Budapest University of Technology and Economics
Műegyetem rkp. 3, Budapest, 1111 HUNGARY
email:*szantai@math.bme.hu*

Edith Kovács
Department of Methodology
Budapest College of Management
Villányi út 11-13, Budapest, 1114 HUNGARY
email:*kovacs.edith@avf.hu*