Acta Universitatis Apulensis,
ISSN 1582-5329, 2011,
Special Issue on Understanding Complex Systems,
Eds. Barna Iantovics, Ladislav Hluchý and Roumen Kountchev

PATTERN RECOGNITION USING t-CHERRY JUNCTION TREE STRUCTURES

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. Our goal here is to develop a pattern recognition algorithm based on an approximation of the joint probability distribution. The approximating joint probability distribution is based on the discovery of some of the conditional independencies underlying the statistical data. For this purpose earlier we introduced the concept of t-cherry junction tree (see [8], [11]). This structure is based on a special kind of graph structure called t-cherry hypertree introduced by J. Bukszár with A. Prékopa and T. Szántai (see [2], [3]). Our method was succesfully tested on a real problem of recognizing Parkinson's disease on the basis of voice disorders (see [9]).

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

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

1. Introduction

Pattern recognition as a field of study developed significantly in the last years. The large number of applications, ranging from the classical ones such as automatic character recognition and medical diagnosis to more recent ones in data mining (such as credit scoring, gene selection, credit card transaction analysis) have increased the researches in this field.

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

The classification uses the Bayes decision rule. The methods using the Bayes decision rule require the knowledge of the multivariate probability distribution function, such as normal distribution whose parameters are estimated from the data. There exist also nonparametric density estimation methods, such as kernel density estimations, which use locally tuned radial basis (e.g. Gaussian) function to interpolate the multi-dimensional density.

In our approach we estimate the underlying multi-variate probability distribution, by the exploitation of the conditional independences between the variables (features), managed by the fitting to the training data a t-cherry junction tree [8], [11]. This represents the novelty of the approach introduced here.

A problem which 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). Our method can be used for discovering those variables (features) which are "informative" (see [13]) for the categorical variable Y.

Throughout this paper we shall use the term of "pattern" to denote a d-dimensional data vector $\mathbf{x} = \left(x^1, \ldots, x^i, \ldots, x^d\right)$ 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 $\{1, \ldots, M\}$ is denoted by Y.

The second section of the paper contains a short review of Markov random fields, t-cherry junction tree, and in this environment we give a new definition of "informative" features. In the third section we describe how the t-cherry junction tree approach can be applied to the classification problem. In the fourth 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. Markov random field, t-cherry junction tree, informative features

Let $X = \{X^i\}_{i=1,\dots,d}$ be a set of discrete finite random variables over the same probability space. Let Λ_i denote the set of values of X^i, and $D = \{1,\dots,d\}$ the set of indeces. In physical literature the set Λ_i is called phase space, and the index set D sometimes denoted by S is called set of sites.

Definition 1.[1] A random field on D with phase space in $\prod_{i=1}^{d} \Lambda_i$ is a collection $X = \{X^i\}_{i=1,\dots,d}$ of random variables X^i with values in Λ_i.

In order to state the Markov property for random fields, we need to introduce the topology on the set of indeces.

Definition 2. [1] A neighborhood system on D is a family $\mathcal{N} = \{N_i\}_{i\in D}$ of subsets of D such that for all $i \in D$

i) $i \notin N_i$

ii) $j \in N_i \Rightarrow i \in N_j$.

Definition 3.[1] The cuple $(D, \mathcal{N})$ is called topology.

Remark 1. If $j \in N_i$ then i can be linked to j by an undirected edge. From this point of view the cuple $(D, \mathcal{N})$ defines an undirected graph, between the indices.

Definition 4. [1] A random field is called Markov random field (MRF) with respect to the neighborhood system $\mathcal{N}$ if $\forall i \in D$, X^i is independent from $X - \{X^j\}_{j\in D\setminus\{N_i\cup\{i\}\}}$ given $\{X^j\}_{j\in N_i}$.

Remark 2. The "neighborhood system $\mathcal{N}$" in Definition 4 can be replaced with the corresponding undirected graph.

Definition 5. For a variable X^i the set $\{X^j\}_{j\in N_i}$ is called set of informative variable.

As an example see the Markov random field of Figure 1. The informative variables for X^6 are X^1, X^3, X^4.

Remark 3. If one is intrested in classification, than it is useful to find the set of informative variables for the classifying variable. For this purpose it is useful to discover the Markov random field, that in practice is mostly unknown.

Throughout the paper we will use the following popular notation:

$$\sum_{\mathbf{x}} P(\mathbf{X}) = \sum_{i_1=1}^{m_1} \dots \sum_{i_d=1}^{m_d} P\left(X^1 = x_{i_1}^1, \dots, X^d = x_{i_d}^d\right).$$

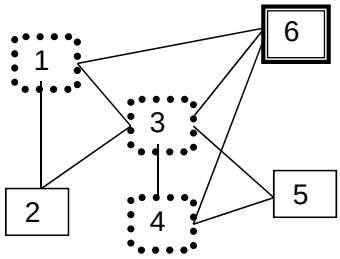

Figure 1: Markov random field over the set of variables with indeces in $D = \{1, 2, \ldots, 6\}$

where $x^k_{i_k}$, $i_k = 1, ..., m_k$ are the possible values of the random variable X^k, $k = 1, ..., d$. Apply similar notation for products, too.

In our earlier papers [8], [11], we have introduced the concept of t-cherry junction tree. This structure is based on the following two definitions:

Definition 6. A tree which fulfills the following properties is called junction tree over **X.**

1) Each node of the tree consists of a subset X_C of X, called *cluster*. To each cluster we assign the joint marginal probability distribution of its random variables.

2) Each edge connecting two clusters of the tree consists of a subset X_S of X given by the intersection of the connected clusters, called *separator*. To each separator we assign the joint marginal probability distribution of its random variables.

3) If two clusters contain a random variable, then all clusters on the path between these two clusters contain this random variable (running intersection property).

4) The union of all clusters is X.

The concept of t-cherrytree hypergraph was introduced by J. Bukszár and A. Prékopa [2] and by J. Bukszár and T. Szántai [3] for constructing sharp lower and upper bounds on the probability of union or intersection of events.

Definition 7. The recursive construction of the k-th order t-cherry tree

(i) The complete graph of (k-1) nodes from D represent the smallest k-th order t-cherry tree.

(ii) By connecting a new vertex from D, with all vertices of $(k-1)$ dimensional complete subgraph of the existing k-th order t-cherry tree, we obtain a new k-th order t-cherry tree.

(iii) Each k-th order t-cherry tree can be obtained from (i) by successive application of (ii).

Remark 4. The k-th order t-cherry tree is a special case of the k-uniform hypergraphs introduced by Tomescu [12].

Definition 8. The set of vertices of the $(k-1)$ dimensional complete subgraph used in step (ii) of Definition 7 is called hyperedge of the k-th order t-cherry tree.

Definition 9. The set of vertices of a hyperedge together with a new vertex is called hypercherry of the k-th order t-cherry tree.

Notation 1 The set of hyperedges of the k-th order t-cherry tree let be denoted by ε_{k-1}.

Notation 2 The set of hypercherries of the k-th order t-cherry tree let be denoted by $\mathcal{C}_k$.

The set of vertices D, the set of hyperedges $\varepsilon_{k-1}$and the set of hypercherries $\mathcal{C}_k$ define the $\Delta_k = (D, \varepsilon_{k-1}, \mathcal{C}_k)$ k-th order t-cherry tree.

Definition 10. The k-th order t-cherry junction tree [11] can be defined in the following way.

1) By using Definition 7 we construct a k-th order t-cherry tree over D : $\Delta_k = (D, \epsilon_{k-1}, \mathcal{C}_k)$.

2) To each hypercherry $(\{i_1, \ldots, i_{k-1}\}, i_k)$ we order a cluster set containing the variables $\left\{X_{i_1}, \ldots, X_{i_{k-1}}, X_{i_k}\right\}$.

3) To each hyperedge $\{i_1, \ldots, i_{k-1}\}$ we oder a separator set containing the variables $\{X_{i_1}, \ldots, \ X_{i_{k-1}}\}$ (edge of the junction tree).

3. The classification method based on t-cherry junction tree

First we introduce some notations and assumptions, see [5].

Let $(\mathbf{X}, Y)^T$ be an $R^d \times \{1, \ldots, M\}$ valued random vector. A classifier is constructed on the basis of a training set $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ and is denoted

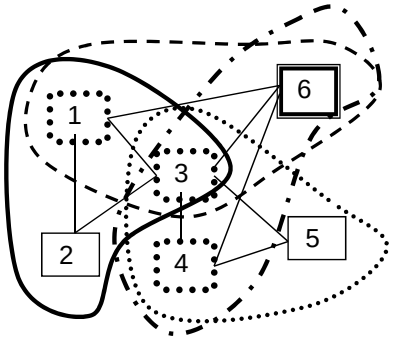

Figure 2: Clusterization of MRF of Figure 1 which leads to the junction tree of Figure 3

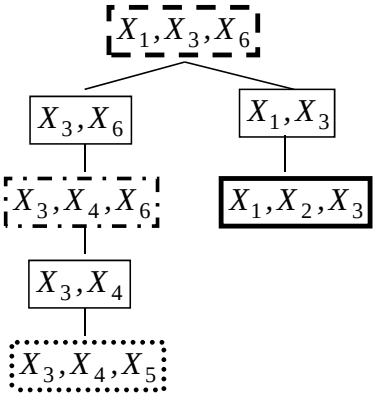

Figure 3: The 3-rd order t-cherry junction tree corresponding to the MRF of Figure 2

by g_n. For a given $\mathbf{x} \in R^d$ the value $y \in \{1, \ldots, M\}$ of Y is guessed by $g_n\left(\mathbf{x}; (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\right)$.

So the classifier g_n is a function:

$$g_n : R^d \times \left\{R^d \times \{1, \ldots, M\}\right\} \longrightarrow \{1, \ldots, M\}.$$

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

We assume that $(\mathbf{X}_1, Y_1)^T, \ldots, (\mathbf{X}_n, Y_n)^T$ 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 (metheorological 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.

The performance of the classifier g_n is measured by the conditional probability error:

$$L_n = L\left(g_n\right) = P\left\{g_n\left(\mathbf{X}; (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\right) \neq Y\right\}.$$

The best possible classifier is defined by:

$$g^* = \underset{g: R^d \to \{1,\ldots,M\}}{\arg\min} P\left(g\left(\mathbf{X}\right) \neq Y\right).$$

We note that g^* depends on the probability distribution of the random vector $(\mathbf{X}, Y)^T$ which usually is unknown.

In the approach presented here we use the class of k-width junction trees (a junction tree which has the largest cluster of a given size k). In this class we search for the best-fitting probability distribution to the training data. The multivariate probability distribution obtained in this way, makes possible to choose the informative features for the classification, and to get a probability distribution containing this features using marginals which contain just k variables. This is one of the novelties provided by our approach.

In order to find the best fitting k-width junction tree, we search in the class of k-width t-cherry junction trees [11].

We intend to find the approximation which minimizes the Kullback-Leibler divergence ([4]):

$$KL = \sum_{\mathbf{x},y} P\left((\mathbf{X}, Y)\right) \log_2 \frac{P\left((\mathbf{X}, Y)\right)}{P_{app}\left((\mathbf{X}, Y)\right)}.$$

From now on the random variable Y will be denoted also by X^{d+1} for notational convenience.

The method for finding a k-th order t-cherry junction tree, using this criterion is presented in [11].

In [11] Theorem 5 claims:

"In the class of k- width junction trees, the k-th order t-cherry junction tree giving the smallest Kullback-Leibler's divergence provides the best approximation of a given $P(\mathbf{X})$ probability distribution'1.

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 the sense (see [11]):

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

where $I\left(X^{i_1},\ldots,X^{i_c}\right)$ is the information content (see [4]) of $P\left(X^{i_1},\ldots,X^{i_c}\right)$.

4. Output : the set of clusters $\mathcal{C}$ and the set of separators $\mathcal{S}$.

5. Select those clusters which contain the variable $X^{d+1}(=Y)$.

6. Select those variables which occur in the clusters selected in step 5 as informative variables.

In Figure 3 the informative clusters for the variable X^6 are (X^1,X^3,X^6) and (X^3,X^4,X^6).

Let us introduce for the set of informative features the notation $X_{\text{info}} = \{X^{i_1},\ldots,X^{i_r}\}$. Let denote $I=\{i_1,\ldots,i_r\}$ the set of indeces. Let be $\mathcal{C}_{\text{info}}$ the set of clusters selected in step 5, and $\mathcal{S}_{\text{info}}$ the set of separators between the clusters selected in step 5. The joint probability distribution of the random vector $\left(\mathbf{X}_{\text{info}},X^{d+1}\right)^T$ let be denoted by $P\left(\left(\mathbf{X}_{\text{info}},X^{d+1}\right)\right)$

$$P_{app}\left(\mathbf{X}_{\text{info}};X^{d+1}\right) = \frac{\prod\limits_{\left(X^{i_1},\ldots,X^{i_c}\right)\subset\mathcal{C}_{\text{info}}} P\left(X^{i_1},\ldots,X^{i_c}\right)}{\prod\limits_{\left(X^{i_1},\ldots,X^{i_s}\right)\subset\mathcal{S}_{\text{info}}} P\left(X^{j_1},\ldots,X^{j_s}\right)^{\left(\nu_{j_1,\ldots,j_s}-1\right)}}, \qquad (1)$$

where $\nu_{j_1,\ldots,j_s} = \#\left\{\{X^{j_1},\ldots,X^{j_s}\} \subset C \mid C \in \mathcal{C}_{\text{info}}\right\}$.

For example in Figure 3:

$$P_{app}\left(X_1,X_3,X_4;X_6\right) = \frac{P\left(X_1,X_3,X_6\right)P\left(X_3,X_4,X_6\right)}{P\left(X_3,X_6\right)},$$

Remark 5. If the number of selected informative variables is not too high one can use their joint marginal probability distribution. This can be determined from the probability distribution underlying the training data. Else it is useful to use formula (1).

We are going to define a classifier. This represents the second novelty of this paper.

In the following we will refer to $P\left(\left(\mathbf{X}_{\text{info}},X^{d+1}\right)\right)$ as $P_{\text{info}}\left(X^{i_1},\ldots,X^{i_r},Y\right)$. This probability is the corresponding marginal determined from the training set or its approximation given by (1).

In order to introduce the classifier we have to make some notations.

We suppose to have a set of training vectors : $\left\{\left(\mathbf{x}_t,y_t\right)\right\}_{t\in\{1,\ldots,n\}}$, where usually $n > \frac{4}{5}N$, where N is the total number of the available observation vectors. These can be selected randomly from the given data set at the beginnig of the analysis. The remaining data constitute the test set.

Let be

$$\begin{aligned}\mathcal{P}_{inf}^{i,k}\left(x_t^{i_1},\ldots,x_t^{i_r}\right) \;=\; & \Big\{P_{\text{info}}\left(X^{j_1}=x_t^{j_1},\ldots,X^{j_k}=x_t^{j_k},Y=i\right)\mid \\ & \left\{X^{j_1},\ldots,X^{j_k}\right\}\subset X_{inf}, \\ & P_{\text{info}}\left(X^{j_1}=x_t^{j_1},\ldots,X^{j_k}=x_t^{j_k},Y=i\right)>0\Big\}\end{aligned}$$

and

$$k^* = \max\left\{k \mid k\in\{1,\ldots,r\};\ \mathcal{P}_{inf}^{i,k}\left(x_t^{i_1},\ldots,x_t^{i_r}\right)\neq\phi\right\}.$$

We define the following random variable:

$$\gamma^{k^*}_{x_t^{i_1},\ldots,x_t^{i_r}}: \quad \begin{pmatrix} 1 & \ldots & i & \ldots & M \\ p_1 & \ldots & p_i & \ldots & p_M \end{pmatrix}$$

where $p_i = \dfrac{\sum\limits_{\mathcal{P}_{inf}^{i,k^*}\left(x_t^{i_1},\ldots,x_t^{i_r}\right)} P_{\text{info}}\left(X^{j_1}=x_t^{j_1},\ldots,X^{j_{k^*}}=x_t^{j_{k^*}},Y=i\right)}{\sum\limits_{\mathcal{P}_{inf}^{i,k^*}\left(x_t^{i_1},\ldots,x_t^{i_r}\right)} P_{\text{info}}\left(X^{j_1}=x_t^{j_1},\ldots,X^{j_{k^*}}=x_t^{j_{k^*}}\right)}, i=1,\ldots,M.$

Remark 6. It is easy to see that $\sum_{i=1}^{M} p_i = 1$.

Definition 11. The classifier $g_{n,\text{info}}^{(r)}$ is defined in the following way:

$$g_{n,\text{info}}^{(r)} : R^d \times \{R^r \times \{1,\ldots,M\}\} \longrightarrow \{1,\ldots,M\}, r < d$$

$$g_{n,\text{info}}^{(r)}\left(x^1,\ldots,x^d\right) = \underset{i\in\{1,\ldots,M\}}{\arg\max}\, P\left(\gamma_{x^{i_1},\ldots,x^{i_r}}^{k^*} = i\right).$$

Remark 7. One can see from this definition that for classifying a new d-dimensional vector we use a number of r informative features only.

Definition 12. The performance of the classifier $g_{n,\text{info}}^{(r)}$ is measured by the conditional probability of error:

$$L_n = L\left(g_{n,\text{info}}^{(r)}(\mathbf{X})\right) = P\left(g_{n,\text{info}}^{(r)}(\mathbf{X}) \neq Y \mid (X_1,Y_1),\ldots,(X_n,Y_n)\right).$$

In practice the probability of error can be approximated by the relative frequency of errors in the set of test vectors.

4. 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 [9]: "Research has shown that approximately 90% of people with PD exhibit some form of vocal impairment [7], [10]. Vocal impairment may also be one of the earliest indicators for the onset of the illness [6]".

The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre 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.

The 23 features are:

MDVP: Fo(Hz) - Average vocal fundamental frequency - X^1

MDVP: Fhi(Hz) - Maximum vocal fundamental frequency - X^2;
MDVP: Flo(Hz) - Minimum vocal fundamental frequency - X^3;
MDVP: Jitter(%)- X^4,
MDVP:Jitter(Abs)- X^5,
MDVP:RAP - X^6,
MDVP:PPQ - X^7,
Jitter:DDP - Several measures of variation in fundamental frequency - X^8
MDVP:Shimmer - X^9,
MDVP: Shimmer(dB) - X^{10},
Shimmer:APQ3 - X^{11},
Shimmer:APQ5 - X^{12},
MDVP:APQ - X^{13},
Shimmer:DDA - Several measures of variation in amplitude- X^{14}
NHR,HNR Two measures of ratio of noise to tonal components in the voice
status - X^{15}, X^{16}
Health status of the subject (one) - Parkinson's, (zero) - healthy - $X^{17} = Y$
RPDE,D2 - Two nonlinear dynamical complexity measures - X^{18}, X^{19}
DFA - Signal fractal scaling exponent -X^{20}
spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation X^{21}, X^{22}, X^{23}.

MDVP stands for denoting measures introduced in the Multi Dimensional Voice Program by Kay Pentax, which became a standard in voice analysis.

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. The 5-th order t-cherry junction tree has the following clusters:

$$
\begin{array}{ccccc}
(X^1, & X^2, & X^3, & X^{10}, & X^{16}) \\
(X^1, & X^3, & X^{10}, & X^{16}, & X^{20}) \\
(X^1, & X^6, & X^{10}, & X^{16}, & X^{20}) \\
(X^1, & X^{10}, & X^{16}, & X^{19}, & X^{20}) \\
(X^1, & X^{10}, & X^{17}, & X^{19}, & X^{20}) \\
(X^4, & X^6, & X^8, & X^{15}, & X^{16}) \\
(X^4, & X^6, & X^{10}, & X^{15}, & X^{16}) \\
(X^4, & X^7, & X^{10}, & X^{13}, & X^{16}) \\
(X^4, & X^{10}, & X^{13}, & X^{15}, & X^{16}) \\
(X^5, & X^{16}, & X^{19}, & X^{21}, & X^{22}) \\
(X^6, & X^{10}, & X^{15}, & X^{16}, & X^{23}) \\
(X^6, & X^{10}, & X^{16}, & X^{20}, & X^{23}) \\
(X^9, & X^{10}, & X^{11}, & X^{12}, & X^{13}) \\
(X^9, & X^{10}, & X^{11}, & X^{13}, & X^{14}) \\
(X^{10}, & X^{11}, & X^{12}, & X^{13}, & X^{16}) \\
(X^{10}, & X^{12}, & X^{13}, & X^{15}, & X^{16}) \\
(X^{10}, & X^{16}, & X^{19}, & X^{20}, & X^{21}) \\
(X^{16}, & X^{19}, & X^{20}, & X^{21}, & X^{22}) \\
(X^{18}, & X^{19}, & X^{20}, & X^{21}, & X^{22})
\end{array}
$$

Only the 5-th cluster is informative $(X^1, X^{10}, X^{17} = Y, X^{19}, X^{20})$. The informative features are: Fo, shimmer, DFA, spread1. Using the probability distribution $P\left(X^1, X^{10}, X^{19}, X^{20}, Y\right)$, we obtained a 93% correct classification performance for the classifier. The authors of paper [9] got 91.4 % correct classification performance using ten features and a Kernel support vector machine.

5. 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 $g^{(r)}_{n,\text{info}}$ 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] P. Bremaud, Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues, Springer, USA, 1999.

[2] J. Bukszár and A. Prékopa, Probability bounds with cherry trees, Mathematics of Operational Research, 26 (2001), 174–192.

[3] J. Bukszár and T. Szántai, Probability bounds given by hypercherry trees, Optimization Methods and Software, 17 (2002), 409–422.

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

[5] L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer, New York, (1996).

[6] J. R. Duffy, Motor Speech Disorders: Substrates Differential Diagnosis and Management, 2nd ed. St Louis, MO: Elsevier, 2005.

[7] 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.

[8] E. Kovács and T. Szántai, On the approximation of discrete multivariate probability distribution using the new concept of t-cherry junction tree, in: Proceedings of the IFIP/IIASA/GAMM-Workshop on "Coping with Uncertainty", held at the International Institute for Systems Analysis (IIASA), Laxenburg, Austria, 10–12, December, 2007. Lecture Notes in Economics and Mathematical Systems, 581, Coping with Uncertainty, Modeling and Policy Issues, Springer, Heidelberg, (2008), pp. 39–56.

[9] 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.

[10] 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.

[11] T. Szántai and E. Kovács, Hypergraphs as mean of discovering the dependence structure of a discrete multivariate probability distribution, in: Proceedings of the APMOD 2008 (APplied mathematical programming and MODelling), Bratislava, Slovakia, 28-31, May 2008, Annals of Operations Research, (2010), to appear.

[12] I. Tomescu, Hypertrees and Bonferroni inequalities, J. Combin. Theory, 41, (1986) 209–217.

[13] I. Guyon, N. Matic, V. Vapnik, Discovering informative patterns and data cleaning, Advances in Knowledge Discovery and Data Mining, AAAI, USA, (1996), 181–203.

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*