Multilabel Structured Output Learning with Random Spanning Trees of Max-Margin Markov Networks (supplementary material) by: Mario Marchand, Hongyu Su, Emilie Morvant, Juho Rousou, and John Shawe-Taylor

Lemma 1

Let $\hat{\mathbf{w}}_T \stackrel{def}{=} \mathbf{w}_T/\|\mathbf{w}_T\|$, $\hat{\boldsymbol{\phi}}_T \stackrel{def}{=} \boldsymbol{\phi}_T/\|\boldsymbol{\phi}_T\|$. *Let* $\mathcal{U}(G)$ *denote the unifor*
we have

$$F(\mathbf{w}, x, \mathbf{y}) = \mathop{\mathbf{E}}_{T\sim\mathcal{U}(G)} a_T \langle \hat{\mathbf{w}}_T, \hat{\boldsymbol{\phi}}_T(x, \mathbf{y})\rangle, \text{ where } a_T$$

$$\mathop{\mathbf{E}}_{T\sim\mathcal{U}(G)} a_T \leq$$

$$\frac{1}{\ell^{-2}} \frac{\ell}{2} \sum_{T\in\mathcal{S}(G)}$$

$$\mathop{\mathbf{E}}_{T\sim\mathcal{U}(G)} \|\mathbf{w}_T\|$$

Multi-task Drug Bioactivity Classification with Graph Labeling Ensembles

Hongyu Su and Juho Rousu

Department of Computer Science,

Finland

.fi

JMLR: Workshop and Conference Proceedings 29:404–418, 2013 ACML 2013

Multilabel Classification through Random Graph Ensembles

Hongyu Su HONGYU.SU@AALTO.FI
Juho Rousu JUHO.ROUSU@AALTO.FI
Helsinki Institute of Information Technology (HIIT)
Department of Information and Computer Science

Structured Output Prediction of Anti-cancer Drug Activity

Hongyu Su, Markus Heinonen, and Juho Rousu

Department of Computer Science
P.O. Box 68, 00014 Universit
{hongyu.su,markus.heinonen,juh
http://www.cs.helsinki

Abstract. We present a structured out
sifying potential anti-cancer drugs. Ou
description of a molecule and predicts t
cer cell lines in one shot. Statistical dep
are encoded by a Markov network that
represent similarity according to an aux
resented via kernels based on molecula
is applied to separate correct multilabe
formance of the multilabel classificatio
ments with NCI-Cancer data containin
of drug-like molecules against 59 cance
our method outperforms the state-of-th

1 Introduction

Machine learning has become increasingly
viable molecular structures are searched
In particular, Quantitative Structure-Activi
lating the molecular structures to bioactiv
toxicity, etc.) are routinely built using stat
ods. In particular, the costly pre-clinical
candidates can be focused to the most pron
models are available [16].

Molecular classification—the task of pre
bioactivity of interest—has been tackled w
inductive logic programming [9] and artificia
decade kernel methods [11,16,4] have emer
way to handle the non-linear properties of c
based methods have obtained promising res
methods focusing on a single target variab
to drug screening applications where large
handled.

In this paper we propose, to our knowle
proach for molecular classification. Our met

T.M.H. Dijkstra et al. (Eds.): PRIB 2010, LNBI 6282,

Structured Prediction of Network Response

Hongyu Su HONGYU.SU@AALTO.FI
Aristides Gionis ARISTIDES.GIONIS@AALTO.FI
Juho Rousu JUHO.ROUSU@AALTO.FI
Helsinki Institute for Information Technology (HIIT)
Department of Information and Computer Science, Aalto University, Finland

Abstract

information propagation, opinion formation, adoption of
echnological innovations, viral marketing, and disease
preading (De Choudhury et al., 2010;
Watts & Dodds, 2007).

nfluence models typically consider *ac*
he network nodes. Examples of such
ng a product or (re)posting a news st
etwork. Often, network nodes perfor
esult of influence from neighbouring
f different models have been propose
nce in a network, most notably the i
nd the linear-threshold models (Kemp
he other hand, performing an action
esult to an external (out of the netw
ation that has also been subject to
s (Anagnostopoulos et al., 2008). A
ade by existing models is that influen
ends only on the nodes that perform t
he action itself.

A central question in the study of netw
nfer the latent structure that governs t
s. This question can be formulated i
ne case no underlying network is ava
ews agencies that do not link each
o infer the hidden network structure,
licit edges between the network no
t al., 2010; Du et al., 2012; Eagle et
Rodriguez et al., 2010; 2011). Howe
n unnecessarily hard one to solve in
On the other hand, in many applicati
nown (e.g., "follower" links in twitte
uestion is to estimate the hidden varia
odel (Goyal et al., 2010; Saito et al.,

The present paper is motivated by the
ion: the influence between two nodes
ot depend only on the nodes and th
lso depends on the action under con
mple, if u and v represent users in t
uenced from u regarding topics relate

Structured Prediction of Network Response Supplementary Material

Hongyu Su HONGYU.SU@AALTO.FI
Aristides Gionis ARISTIDES.GIONIS@AALTO.FI
JUHO.ROUSU@AALTO.FI
ersity, Finland

-hard problem.

as searching for an optimal configuration of node
$\{y_e \in \{\texttt{nn}, \texttt{pn}, \texttt{pp}\}\}_{e\in E}$. For each edge e and its
find a set of node labels that maximizes the sum

AX-CUT problem (Garey & Johnson, 1979), where
raph into two parts so that the number of edges
hard.

ce problem. Given an instance of MAX-CUT, i.e.,
oblem as follows. We first create a directed graph
.e., for each $(u, v) \in E$ we add (u, v) and (v, u) in
$e) = s_{\texttt{nn}}(e) = 0$ and $s_{\texttt{pn}}(e) = 1$. In addition, the
. It is not difficult to see that the reduction only
MAX-CUT problem is also optimal to this special
□

ence algortihm

m over the scores of consistent edge labels. To
and labels. We show in Eq. 1 that the inference
l vertices and the associated scores by initializing
of the inference problem enables us to design an
he network.

ithm 1. The algorithm starts with an activated
also maintains a priority list of vertices by their
gorithm pops from the top of the priority list the

(v_i).

ores $F_m(v)$ and positions in the list (line 6). The

Multilabel Structured Output Learning with Random Spanning Trees of Max-Margin Markov Networks

Mario Marchand
Département d'informatique et génie logiciel
Université Laval
Québec (QC), Canada
mario.marchand@ift.ulaval.ca

Hongyu Su
Helsinki Institute for Information Technology
Dept of Information and Computer Science
Aalto University, Finland
hongyu.su@aalto.fi

Emilie Morvant*
LaHC, UMR CNRS 5516
Univ. of St-Etienne, France
emilie.morvant@univ-st-etienne.fr

Juho Rousu
Helsinki Institute for Information Technology
Dept of Information and Computer Science
Aalto University, Finland
juho.rousu@aalto.fi

John Shawe-Taylor
Department of Computer Science
University College London
London, UK
j.shawe-taylor@ucl.ac.uk

Abstract

We show that the usual score function for conditional Markov networks can be written as the expectation over the scores of their spanning trees. We also show that a small random sample of these output trees can attain a significant fraction of the margin obtained by the complete graph and we provide conditions under which we can perform tractable inference. The experimental results confirm that practical learning is scalable to realistic datasets using this approach.

1 Introduction

Finding an hyperplane that minimizes the number of misclassifications is $\mathcal{NP}$-hard. But the support vector machine (SVM) substitutes the hinge for the discrete loss and, modulo a margin assumption, can nonetheless efficiently find a hyperplane with a guarantee of good generalization. This paper investigates whether the problem of inference over a complete graph in structured output prediction can be avoided in an analogous way based on a margin assumption.

We first show that the score function for the complete output graph can be expressed as the expectation over the scores of random spanning trees. A sampling result then shows that a small random sample of these output trees can attain a significant fraction of the margin obtained by the complete graph. Together with a generalization bound for the sample of trees, this shows that we can obtain good generalization using the average scores of a sample of trees in place of the complete graph. We have thus reduced the intractable inference problem to a convex optimization not dissimilar to a SVM. The key inference problem to enable learning with this ensemble now becomes finding the maximum violator for the (finite sample) average tree score. We then provide the conditions under which the inference problem is tractable. Experimental results confirm this prediction and show that

*Most of this work was carried out while E. Morvant was affiliated with IST Austria, Klosterneurburg.

Mach Learn
DOI 10.1007/s10994-014-5465-9

Multilabel classification through random graph ensembles

Hongyu Su · Juho Rousu

Received: 16 January 2014 / Accepted: 14 July 2014

Abstract We present new methods for multilabel classification, relying on ensemble learning on a collection of random output graphs imposed on the multilabel, and a kernel-based structured output learner as the base classifier. For ensemble learning, differences among the output graphs provide the required base classifier diversity and lead to improved performance in the increasing size of the ensemble. We study different methods of forming the ensemble prediction, including majority voting and two methods that perform inferences over the graph structures before or after combining the base models into the ensemble. We put forward a theoretical explanation of the behaviour of multilabel ensembles in terms of the diversity and coherence of microlabel predictions, generalizing previous work on single target ensembles. We compare our methods on a set of heterogeneous multilabel benchmark problems against the state-of-the-art machine learning approaches, including multilabel AdaBoost, convex multitask feature learning, as well as single target learning approaches represented by Bagging and SVM. In our experiments, the random graph ensembles are very competitive and robust, ranking first or second on most of the datasets. Overall, our results show that our proposed random graph ensembles are viable alternatives to flat multilabel and multitask learners.

Keywords Multilabel classification · Structured output · Ensemble methods · Kernel methods · Graphical models

1 Introduction

Multilabel and multitask classification rely on representations and learning methods that allow us to leverage the dependencies between the different labels. When such dependencies

Editors: Cheng Soon Ong, Tu Bao Ho, Wray Buntine, Bob Williamson, and Masashi Sugiyama.

H. Su (✉) · J. Rousu
Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University, Konemiehentie 2, 02150 Espoo, Finland
e-mail: hongyu.su@aalto.fi

J. Rousu
e-mail: juho.rousu@aalto.fi

Published online: 15 August 2014 Springer