Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Neurocomputing 64 (2005) 93–105
NEUROCOMPUTING
www.elsevier.com/locate/neucom

SVM learning with the Schur–Hadamard inner product for graphs

Brijnesh J. Jain*, Peter Geibel, Fritz Wysotzki

Department of Computer Science, Technical University Berlin, Methods of Artificial Intelligence, Sekr. Fr 5-8, Faculty IV, TU Berlin, Franklinstr. 28/29, D-10587 Berlin, Germany

Available online 23 January 2005

Abstract

We apply support vector learning to attributed graphs where the kernel matrices are based on approximations of the Schur–Hadamard inner product. The evaluation of the Schur–Hadamard inner product for a pair of graphs requires the determination of an optimal match between their nodes and edges. It is therefore efficiently approximated by means of recurrent neural networks. The optimal mapping involved allows a direct understanding of the similarity or dissimilarity of the two graphs considered. We present and discuss experimental results of different classifiers constructed by a SVM operating on positive semi-definite (psd) and non-psd kernel matrices.

Keywords: Support vector machine; Graph kernel; Hopfield network; Recurrent network; Schur–Hadamard product

1. Introduction

In many application areas, support vector machines (SVMs, [15,13]) yield excellent learning results. SVMs are linear learning machines. If the non-linear discrimination of classes is required, the so-called *kernel trick* is used. Using a kernel

*Corresponding author. Tel.: +49 30 314 23938; fax: 49 30 314 24913.
E-mail address: bjj@cs.tu-berlin.de (B.J. Jain).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.11.011

for learning corresponds to mapping the original input space to a feature space often having a higher dimension.

So far, most research on kernel methods has focused on learning from attribute value data. The investigation on kernel methods for attributed graphs, however, has recently started [2] and is still widely unexplored though graphs are a more adequate representation of patterns in structured domains than feature vectors.

Examples are the classification of chemical compounds (e.g. [14]), and the classification of images. In the case of chemical compounds, graphs provide a natural representation of the data. In order to classify images—a typical pattern recognition task—each image is described by its line or region adjacency graph. Learning on such a graphical representation yields classifiers that are invariant with respect of translations and rotations.

By providing an appropriate kernel, a given domain of graphs is embedded into a linear space thus allowing the application of kernel-based methods like the SVM.

Since the kernel function defines the inner product in the feature space, it can be regarded as a *similarity measure* for graphs. In this article, we use a graph similarity that is based on the Schur–Hadamard inner product adapted for graphs. The optimal mapping involved allows a direct understanding of the similarity or dissimilarity of the two graphs considered.

The computation of the Schur–Hadamard inner product requires the determination of optimal permutations for pairs of graphs. In order to solve this NP-complete task, a Hopfield-type network is employed. Thus, this approach combines recurrent neural networks with SVMs.

Despite the fact that approximations of the Schur–Hadamard inner product is not positive definite in general, we will show experimentally that nevertheless very good classification results can be obtained using the SVMLight [10] extended with the respective *pseudo-graph kernels*. Additionally, we investigate the usage of similarity vectors constructed from the Schur–Hadamard inner product for learning as proposed in [6].

This article is structured as follows. Section 2 is a collection of basic notations and definitions and discusses kernel functions for graphs. In Section 3, we define the Schur–Hadamard inner product of graphs and further propose a neural network approach to approximate the Schur–Hadamard inner product. In Section 4, we present experiments for a molecule classification task, and for the recognition of digits and letters. Section 5 concludes the article.

2. SVM

SVMs [15,13] have proven to be widely applicable and successful in data classification. Given a set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\} \subseteq \mathscr{X}$ of training objects from a domain $\mathscr{X}$ with corresponding labels $\{y_1, \ldots, y_M\} \subseteq \{+1, -1\}^M$, a SVM learns an optimal hyperplane to separate the training objects in a feature space $\mathscr{F}$ by solving the