Novel kernels for error-tolerant graph classification

MICHEL NEUHAUS, KASPAR RIESEN* and HORST BUNKE
Institute of Computer Science, University of Bern, Neubrückstrasse 10, CH-3012 Bern, Switzerland

Received 10 April 2007; accepted 29 December 2008

Abstract—One of the major difficulties in graph classification is the lack of mathematical structure in the space of graphs. The use of kernel machines allows us to overcome this fundamental limitation in an elegant manner by addressing the pattern recognition problem in an implicitly existing feature vector space instead of the original space of graphs. In this paper we propose three novel error-tolerant graph kernels — a diffusion kernel, a convolution kernel, and a random walk kernel. The kernels are closely related to one of the most flexible graph matching methods, graph edit distance. Consequently, our kernels are applicable to virtually any kind of graph. They also show a high degree of robustness against various types of distortion. In an experimental evaluation involving the classification of line drawings, images, diatoms, fingerprints, and molecules, we demonstrate the superior performance of the proposed kernels in conjunction with support vector machines over a standard nearest-neighbor reference method and several other graph kernels including a standard random walk kernel.

Keywords: Graph classification; error-tolerant graph matching; graph edit distance; graph kernel.

1. INTRODUCTION

Graphs are a popular concept for the representation of structured information in pattern recognition (Bunke, 2000). The main advantage of graphs over the feature vector approach, where each pattern is represented by a vector of fixed dimension, is that graphs allow for the representation of sets of entities by nodes and relations between these entities by edges. The number of nodes and edges in a graph can be adapted to the complexity of the pattern to be represented, so that complex objects are represented by large graphs and simple objects by small graphs. In the general case, nodes and edges can additionally be labeled with attributes. While the representational power of graphs is clearly higher than that of feature vectors, the space of graphs contains a priori almost no mathematical structure, in contrast to feature vector spaces. Basic mathematical operations that can easily be defined in vector spaces, such as the Euclidean distance or linear combinations of patterns, are

*To whom correspondence should be addressed. E-mail: riesen@iam.unibe.ch
not directly available for graphs. Instead, the application-specific definition of graph operations often involves a tedious and time-consuming development process.

Kernel machines offer an elegant solution to the fundamental problem of lack of mathematical structure in the space of graphs. Kernel machines are a class of algorithms for pattern analysis and classification (Schölkopf and Smola, 2002; Shawe-Taylor and Cristianini, 2004). The basic idea of kernel machines is to address a pattern recognition problem in a related vector space instead of the original pattern space. That is, rather than defining mathematical operations in the space of graphs, graphs are mapped into a vector space where these operations are readily available. A key result from kernel function theory states that an explicit mapping is in fact not even required. Instead, for kernel machines to be applicable, it is sufficient to define kernel functions that extract the information from the space of patterns that is relevant for recognition, which can be achieved through an implicit embedding of the pattern space into a vector space. Accordingly, graph kernel functions render a large class of statistical pattern recognition algorithms applicable to graphs. A number of kernel functions have been proposed for graphs (Gärtner, 2008; Jain et al., 2005; Kashima et al., 2003), but these kernels are to a large extent applicable to unlabeled graphs only or unable to deal sufficiently well with strongly distorted data.

Graph matching for structural pattern recognition has been a focus of intensive research for many years and has lead to a large variety of matching methods including relaxation labeling (Wilson and Hancock, 1997), recursive neural networks (Frasconi et al., 1998), genetic algorithms (Cross et al., 1997), spectral decomposition of graphs (Caelli and Kosinov, 2004; Umeyama, 1988), or random walks (Gori et al., 2005). For an extensive review of graph matching methods we refer to Conte et al. (2004). A method that is recognized as one of the most flexible error-tolerant graph matching techniques is graph edit distance (Bunke and Allermann, 1983; Sanfeliu and Fu, 1983). Graph edit distance is applicable to arbitrarily structured graphs and any kind of node and edge labels. The basic idea is to define the dissimilarity of two graphs by evaluating how expensive it is to transform one graph into the other one. The edit distance concept is not constrained to graphs, but can also be used for the matching of strings, trees, and hypergraphs and can thus be considered a universal structural matching paradigm. In the present paper, we propose three novel error-tolerant graph kernels that are to a certain extent related to graph edit distance. The rationale for this is to combine the flexibility of graph edit distance for graph matching with the power of kernel machines for classification.

2. GRAPH EDIT DISTANCE

Graph edit distance is an error-tolerant dissimilarity measure for graphs. The edit distance concept has originally been established in the context of the string-to-string correction problem (Wagner and Fischer, 1974) and subsequently been adapted from strings to trees (Selkow, 1977) and graphs (Bunke and Allermann, 1983;