A Combined Dimensional Kernel Method for Graph Classification

Tiejun Cao, School of Electrical and Information Engineering, Hunan International Economics University, Changsha, China

ABSTRACT

The data containing structural information is an important problem in the field of machine learning. Kernel methods is an effective technique for solving such problems. A combined dimension kernel method is proposed or graph classification in this paper. A two-dimensional kernel is first constructed in this method, and it incorporates one-dimensional information to characterize the molecular chemistry, and then a three-dimensional kernel is constructed based on the knowledge of molecular mechanics to characterize the physical properties of the molecule. On this basis, the kernel of different dimensions is integrated, and the quadratic programming problem with quadratic constraints is solved to obtain the optimal kernel combination. The experimental results show that the proposed method has better performance than the prior technology, and it outperforms the existing algorithm.

KEYWORDS

Ensemble Learning, Graph Classification, Kernel Method, Machine Learning, Structure Information

1. INTRODUCTION

Many of the commonly used classification algorithms assume that the sample is represented by a vector, whereas the sample is usually structured in many practical applications, such as biology RNA, DNA sequence representation (Durbin, Eddy, Krogh et al., 1998; Zeng, 2015; Gao, Cheng et al., 2016), natural language processing tree representation (Manning & Schütze, 1999), XML semi-structured representation (Abiteboul, Buneman, & Suciu, 2000) and chemical molecules map representation (Swamidass, Chen, Bruand et al., 2005; Xu, Peng et al., 2016) and so on. How to use the structural information contained in these samples to effectively learn, is an important problem in the field of machine learning. The results of this work play a role in toxicity detection, mutagenesis, carcinogenic detection of chemical molecules and many other applications (Kramer & De Raedt, 2001; Inokuchi, Washio, & Motoda, 2000; Wang, Qin et al., 2012).

The kernel method can guarantee the generalization performance by mapping the data into the high-dimensional feature space and optimizing the structural risk in the feature space. Because this method only needs to construct a kernel function to measure the similarity between the samples, and it is not limited by the specific representation of the sample, it can effectively learn structural data such as graph (Scholkopf & Smola, 2002). A suitable kernel function is constructed for a given learning task, it is the core problem of the kernel approach. The difficulty lies in the need to efficiently and rationally use the sample information.
to describe the similarity of samples, but also to meet the semi - positive definite of the matrix and to get the optimal solution. As far as the structure of chemical molecules is concerned, many researchers have proposed a variety of kernel functions, such as random path kernel (Gartner, Flach, & Wrobel, 2003), best matching kernel (Frohlich, Wegner, Sieker et al., 2005), interval kernel (Kashima, Tsuda, & Inokuchi, 2003), and so on. Recently, Swamidass et al. had designed a family of kernel functions for one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) representations of chemical molecules, respectively, with knowledge domain. However, this method only considers the chemical characteristics of the molecule, it ignored the physical characteristics of the molecule (Swamidass, Chen, Bruand et al., 2005).

Quadratic Constrained Quadratic Programming (QCQP) is introduced in this paper to generate the optimal combination of 2D and 3D kernels. Experiments show that the proposed method has better performance than the existing methods on the 10 data sets of PTC and NCI, and the performance of Mutag data sets is equal to that of the existing algorithms.

2. RESEARCH BACKGROUND

2.1. Graph Classification Method Based on Kernel Function

Real-world objects usually contain some structural information, so the study of structured data is more and more concerned by the machine learning community. In the early work, the researchers focused on some special cases of graphs, such as string, tree, etc., and they proposed a variety of effective methods, such as spectral kernel (Leslie, Eskin & Noble, 2002), convolution kernel (Collins & Duffy, 2002), such technology is called a semantic-based kernel approach (Gartner, 2003). In 2003, Gartner studied the general graph classification (Gartner, Flach & Wrobel, 2003), and pointed out that the perfect graph kernel (i.e., the feature map is a bijective function) is a problem that is at least as hard as the graph isomorphism, the inner product of graph space will be a NP difficult problem. This means that for the general graph classification problem, the research goal should be to find a good approximation solution, rather than the ideal solution. Using stochastic path approximation data distribution, Gartner proposed a stochastic path algorithm with polynomial time cost, which can effectively describe the similarity between graphs. Since then, many researchers have studied general graph classification and proposed many kernel methods, such as interval kernel (Kashima, Tsuda, & Inokuchi, 2003), Fisher kernel (Jaakkola & Haussler, 1999), diffusion kernel (Kondor & Lafferty, 2002), which are called a model-based kernel approach (Gartner, 2003). In addition, some researchers have proposed the shortest path kernel (Borgwardt & Kriegel, 2005), the best matching kernel (Frohlich, Wegner, Sieker et al., 2005), the ring kernel (Horvath, Gartner & Wrobel, 2004) and so on by using the method of graph theory.

It is worth noting that the graph data of the real problem usually has a bright background, if the domain knowledge can be effectively used, it is possible to obtain better performance. Note that chemical molecules have three dimensional representations, one dimension is a string, two dimensions is a marker map, and three dimensions is a spatial three-dimensional model. Swamidass et al. recently combined domain knowledge to define the corresponding kernel function with three dimensions, which can obtain good experimental results with a small calculation overhead (Swamidass, Chen, Bruand et al., 2005).

However, by studying the work of Swamidass and others, we found that although Swamidass et al.’s 3D kernel had the most abundant information, it did not perform as well. The reason may be that the 1D, 2D and 3D kernels which were constructed by Swamidass and others are independent of
A Classical Uncertainty Principle for Organizations

www.igi-global.com/chapter/classical-uncertainty-principle-organizations/13625?camid=4v1a