Pattern Recognition 73 (2018) 89–98

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

A hybrid reproducing graph kernel based on information entropy

Lixiang Xu[a,b,c], Xiaoyi Jiang[c], Lu Bai[d], Jin Xiao[e], Bin Luo[a,*]

[a] School of Computer Science and Technology, Anhui University, Hefei 230601, Anhui, PR China
[b] Department of Mathematics and Physics, Hefei University, Hefei 230601, Anhui, PR China
[c] Faculty of Mathematics and Computer Science, University of Münster, Einsteinstrasse 62, 48149 Münster, Germany
[d] School of Information, Central University of Finance and Economics, Beijing, 100081, PR China
[e] Business School, Sichuan University, Chengdu 610064, Sichuan, PR China

ARTICLE INFO

Article history:
Received 18 August 2016
Revised 18 June 2017
Accepted 26 July 2017
Available online 27 July 2017

Keywords:
Differential operator
Reproducing kernel
Information entropy
Graph kernel
Graph classification

ABSTRACT

A number of graph kernel-based methods have been developed with great success in many fields, but very little research has been published that is concerned with a graph kernel in Reproducing Kernel Hilbert Space (RKHS). In this paper, we firstly start with a derived expression for two forms of information entropy of an undirected graph. They are approximated von Neumann entropy and Shannon entropy, and depend on vertex degree statistics. Secondly, we show the basic solution of a generalized differential operator. This solution is a specific reproducing kernel called the H^1-reproducing kernel in H^1-space, and then it is proven to satisfy the condition of Mercer kernel. Thirdly, based on the two aforementioned forms of information entropy and H^1-reproducing kernel, we define two reproducing graph kernels: one is approximated von Neumann entropy reproducing graph kernel (AVNERGK), the other is Shannon entropy reproducing graph kernel (SERGK). And then we prove that they satisfy the condition of Mercer kernel. Finally, to obtain better classification results, we further propose a hybrid reproducing graph kernel (HRGK) based on the two reproducing graph kernels. We use the HRGK as a means to establish the similarity between a pair of graphs. Experimental results reveal that our method gives better classification performance on graphs extracted from several graph datasets.

1. Introduction

There has recently been an increasing interest in learning and mining data using graph structures. To analyze and understand these data, we need some graph similarity measure methods that can recognize structural graph datasets. Important graph-related similarity measure methods include graph isomorphism testing [24], graph matching [12,18,21], graph clustering [23], graph edit distance [20,22], and median graph computation [15], and so on. Graph kernels are powerful tools for structural analysis in machine learning. The main advantage of using graph kernels is that they provide an implicit embedding of graphs in a high dimensional space where structural information is better preserved [3,32].

1.1. Related work

Most of the existing graph kernels are instances of the R-convolution kernels proposed by Haussler [34]. Here graph kernels are computed by comparing the similarity of each of the decompositions of the two graphs. Depending on how the graphs are decomposed, we obtain different kernels. The R-convolution kernels can be categorized into the following classes, namely graph kernels based on comparing all pairs of a) walks [14], b) paths [7], c) cycles [2], and d) subgraph or subtree structures [4,27]. However, the R-convolution kernels only count the number of pairwise isomorphic substructures. As a result, the substructures having no corresponding isomorphic substructures are discarded. Moreover, the R-convolution kernels are more computationally expensive, and do not easily scale up to large sized structures. To overcome the shortcomings of existing R-convolution kernels, most existing graph kernels compromise to use substructures of limited sizes, and examples include a) the shortest path graph kernel [7], b) the graphlet count graph kernel [27], c) the fast neighborhood subgraph pairwise distance kernel [9], and d) the backtrackless kernel [2].

In recent years, a number of state-of-the-art graph kernel methods have also been reported. For example, to implement the computation of the von Neumann entropy more efficiently, Han et al. [13] have introduced how the computation can be implemented quadratically in the number of vertices [36]. To extend this work further, Ye et al. [37] have developed how the von

* Corresponding author.
E-mail addresses: xulixianghf@163.com (L. Xu), luobin@ahu.edu.cn, luobinahu@163.com (B. Luo).

http://dx.doi.org/10.1016/j.patcog.2017.07.025
0031-3203/

Neumann entropy for undirected graphs can be extended to directed graphs. Furthermore, Bai et al. [3] investigated whether the Jensen–Shannon divergence can be used as a means of establishing a graph kernel, and proposed the Jensen–Shannon graph kernel. To measure the isomorphisms between hypergraphs straightforwardly, they then built on this initial work and presented a hypergraph kernel computed using substructure isomorphism tests [5]. Later, they [6] used the quantum Jensen–Shannon divergence as a means of measuring the information theoretic dissimilarity of graphs and thus developed a novel quantum Jensen–Shannon graph kernel. Moreover, Xu et al. [32] proposed a mixed Weisfeiler–Lehman (WL) graph kernel framework based on a family of efficient WL graph kernels. It can be competitive with or outperforms the corresponding single WL graph kernel in the experiments. Based on the latest advancements in language modeling and deep learning, Pinar et al. [38] have presented Deep Graph Kernels, which is a unified framework to learn latent representations of sub-structures for graphs. Zhao et al. [39] have derived a graph kernel to quickly and accurately compute the graph similarity in videos for automatic behavior analysis. Their approach can be plugged into any kernel-based classifier.

There have been many other successful attempts to classify or cluster graphs using graph kernels, such as edge kernels [30], Weisfeiler–Lehman graph kernels [26], fast multi-view segment graph kernels [33], and many others. However, it should firstly be noted that most of the graph kernel methods mentioned above tend to be burdensome and have a high time complexity. The runtime may be several days for graphs of large sizes (e.g. a graph having more than one thousand vertices). Secondly, some kernels may not obtain better experiment results.

1.2. Contributions and paper organization

In this paper, we explore how a simplification can be used to efficiently compute the graph kernel between graphs. The resulting computations depend on the node degree distribution over the graph and reproducing kernel function, and can be simply computed for both the original graphs. The main contributions of this paper are as follows. Firstly, the reproducing kernel function possesses many nice properties, such as the odd order vanishing moment, symmetry and regularity. We therefore show the basic solution of a generalized differential operator by the delta function, and this solution was proven to be a specific H^1-reproducing kernel. We further prove that the H^1-reproducing kernel satisfies the condition of Mercer kernel. Secondly, to develop a novel graph kernel using the divergence measure, based on the approximated von Neumann entropy and Shannon entropy measure for each graph and H^1-reproducing kernel, we define two novel graph kernels based on information entropies. They are the approximated von Neumann entropy reproducing graph kernel (AVNERGK) and Shannon entropy reproducing graph kernel (SERGK). Thirdly, to take into account the structural information which was overlooked in the methods discussed previously and improve the classification accuracy of reproducing graph kennel methods, we propose a hybrid reproducing graph kernel (HRGK) based on two reproducing graph kernels for unattributed graphs. Experiments demonstrate that it is consistently comparable or superior to a number of other existing state-of-the-art graph kernels in terms of the classification accuracy and runtime. Note that this paper is partly motivated by defining two novel reproducing graph kernels (AVNERGK and SERGK). As a second motivation, we intend to mix the two novel reproducing graph kernels to improve the classification performance of graph kernel methods based on reproducing kernel. In future, we shall further discuss multiple graph-kernel learning which are based on the graph kernels of different types, possibly with better diversity.

The remainder of this article is organized as follows. Section 2 describes fundamentals, including the reproducing kernel, conditions of Mercer kernel, node degree distribution, and information entropy for an unattributed graph. In Section 3 we present the reproducing graph kernels AVNERGK and SERGK, and the hybrid reproducing graph kernel. Section 4 provides the experimental evaluation. Finally, a summary in Section 5 concludes the paper.

2. Fundamentals

2.1. Reproducing kernel

A Hilbert Space is an inner product space that is complete and separable with respect to the norm defined by the inner product. A Hilbert space of complex-valued functions which possesses a reproducing kernel is called a RKHS or a proper Hilbert space [41].

Definition 1. A function: K: $E \times E \to C$, $(s, t) \mapsto K(s, t)$ is a reproducing kernel of the Hilbert space H if and only if

(i) $\forall t \in E$, $K(., t) \in H$;

(ii) $\forall t \in E$, $\forall \phi \in H$, $\langle \phi, K(., t) \rangle = \phi(t)$.

This last condition (*ii*) is called the reproducing property [41]: the value of the function ϕ at the point t is reproduced by the inner product of ϕ with $K(., t)$.

2.2. Conditions of Mercer Kernel

If a function satisfies the condition of Mercer kernel, it is the allowable kernel function. For the translation invariant function, we can give the condition of translation invariant kernel function [40,42].

Lemma 1. *The translation invariant function $k(x, x') = k(x - x')$ is an allowable kernel function if and only if the Fourier transform of $k(x)$ satisfies the condition*

$$\hat{k}(w) = (2\pi)^{-\frac{d}{2}} \int_R exp(-jwx)k(x)dx \geq 0 \tag{1}$$

Lemma 1 proposes a simple method to build the kernel function.

The kernel function plays a crucial role in understanding Support Vector Machines (SVMs) and will be an important theme throughout this paper. We shall examine its properties, related algorithms and applications in general pattern analysis.

2.3. Node degree distribution

In order to understand the principle of AVNERGK, we need to make clear the concepts below. Here, we use node degree distribution to calculate the approximated von Neumann entropy. To begin, we denote the graph as $G = (V, E)$ where V is the set of nodes and $E \subseteq V \times V$ is the set of edges. For the adjacency matrix A of graph G, it has elements

$$A(u, v) = \begin{cases} 1 & if\ (u, v) \in E, \\ 0 & otherwise. \end{cases} \tag{2}$$

The degree matrix of graph G is a diagonal matrix D. Its diagonal elements are the node degrees: $D(u, u) = d_u = \sum_{v \in V} A(u, v)$. By the adjacency matrix and the degree matrix, we have the Laplacian matrix $L = D - A$, i.e. the degree matrix minus the adjacency matrix. Further, we obtain the elements of the Laplacian matrix are

$$L(u, v) = \begin{cases} d_v & if\ u = v, \\ -1 & if\ (u, v) \in E, \\ 0 & otherwise \end{cases} \tag{3}$$