Computer Science Review 27 (2018) 88–111

Contents lists available at ScienceDirect

Computer Science Review

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

Review article

The journey of graph kernels through two decades

Swarnendu Ghosh [a], Nibaran Das [a,*], Teresa Gonçalves [b], Paulo Quaresma [b], Mahantapas Kundu [a]

[a] Department of Computer Science, Jadavpur University, India
[b] Department of Informatics, University of Evora, Portugal

ARTICLE INFO

Article history:
Received 11 July 2017
Received in revised form 28 November 2017
Accepted 29 November 2017

Keywords:
Graph kernels
Support vector machines
Graph similarity
Isomorphism

ABSTRACT

In the real world all events are connected. There is a hidden network of dependencies that governs behavior of natural processes. Without much argument it can be said that, of all the known data-structures, graphs are naturally suitable to model such information. But to learn to use graph data structure is a tedious job as most operations on graphs are computationally expensive, so exploring fast machine learning techniques for graph data has been an active area of research and a family of algorithms called kernel based approaches has been famous among researchers of the machine learning domain. With the help of support vector machines, kernel based methods work very well for learning with Gaussian processes. In this survey we will explore various kernels that operate on graph representations. Starting from the basics of kernel based learning we will travel through the history of graph kernels from its first appearance to discussion of current state of the art techniques in practice.

Contents

* Corresponding author.
E-mail addresses: swarbir@gmail.com (S. Ghosh), nibaran.das@jadavpuruniversity.in (N. Das), tcg@di.uevora.pt (T. Gonçalves), pq@di.uevora.pt (P. Quaresma), mahantapas.kundu@jadavpuruniversity.in (M. Kundu).

https://doi.org/10.1016/j.cosrev.2017.11.002
1574-0137/

1. Introduction

Information has always been in the primary focus of researchers in the field of computer science. In our world, most of the available information is represented as networks of meaningfully connected data elements. These connections can signify some sort of interdependence or portray some contextual significance. This relational aspect of information is one of the main challenges for researchers. In this survey will be explored the utility of various graph kernels in this domain of relational information, but before we move on to the details of graph kernels, let us first understand the importance of "graphs" and "kernels" in the field of artificial intelligence. One of the primary tasks is sensible representation of such relational data, so that they could be used to perform machine learning tasks such as classifications, sequence predictions, density estimations and so on.

Information is mainly stored using data structures for computers to process them. While there are many data structures available, the most generic format is a graph. All other data structures are simply some sort of specializations of a graph. As we know, graphs are characterized by their network of nodes connected by edges. Similarly, natural information in general can be broken down to smaller elements that can have some sort of semantic connection hence, this property of graph makes it most suitable for representing relational information. So, the first step of graph based learning is to actually represent the information in the form of a graph. Once that is done the second step is the learning part.

The most straightforward technique for learning is to extract meaningful features from a sample that uniquely predicts its nature. However, that is not always feasible given the dynamic nature of real world problems. Problems can be so complicated that manually extracting features can be really hectic and sometimes humanely impossible. Data in its raw form is not suitable for computational operations. A consistent input space is needed to represent the data in its actual form. The key idea behind finding features is to move the sample from the input space to another dimension where similar samples will be mapped in close proximity while distance between dissimilar samples will be significantly higher. Another branch of machine learning, namely kernel based learning, views the problem from a different perspective. If we can find some metric to map this similarity between samples we can directly map them onto the feature dimension without actually having to learn the features themselves. Another way to explain this is to approximate the nature of the probability distribution of the real world process, also know as the Gaussian process, so that the similar samples stay in close proximity and vice versa. This new dimension is also called an Hilbert space. The entire goal of kernel based learning is to map the available sample space into a suitable Hilbert space. Once we know the Gaussian distribution, also termed as the posterior, it will be much easier to calculate the similarity among samples. Machine learning dived into a new paradigm through the introduction of a special function referred to as a kernel function which can directly map the input space to such feature dimensions. Throughout the next chapters, we will look into details regarding definitions, mathematical concepts and old and modern research works surrounding the application of kernels to the field of graph theory.

As we finish the introductory section we will find our motivation to study more about this domain in next section. Section 3 introduces us to the preliminary concepts of some Gaussian Processes, Kernel based Machine Learning, and Graph Theory. This is absolutely necessary for understanding the concepts of various graph kernels. As we move on to the fourth section, we will discuss the core concepts of graph kernels, starting from the earliest point in the history of research where the first idea of structural kernels was conceived and slowly moving through time to finally analyze a couple of state of the art technologies. Utmost effort has been made to keep all explanation as simple as possible while maintaining enough mathematical formulation to ensure logical clarity.

2. Motivation

Graphs provide one the most generic data structures for representing information. Philosophically speaking a graph represents a network of relationships among objects. All real world phenomena can be interpreted as a system with various components that work in tandem. These relations and interdependence connect these components to form a complex network. Another interpretation may be all real world objects or events can either be described as a network or can be considered to be a part of a larger network. Philosophical arguments have been made in favor of graphs as the most ideal data structure to represent the world in the language of mathematics [1].

In computational terms it has already been mentioned that graph are the most generic form of data structure as all common datatypes can simply be referred to as an instance of a graph. For example, a scalar or a constant can be treated as single node graph, and array or matrix can be seen as a graph where each nodes represent an index in the array and their adjacency is represented by an edge. Stacks and queues have similar structure but with limitation of insertion and deletion property of the nodes. A time series can be modeled by representing time stamps as nodes and connecting each stamp with an edge to the next one.

So, with all this said, the real question is why graphs are not being used as the most common data structure for decades? The simple answer is that handling graphs is complicated. On one hand