Pattern Recognition 48 (2015) 291–301

Contents lists available at ScienceDirect

Pattern Recognition

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

A long trip in the charming world of graphs for Pattern Recognition

Mario Vento

Dept. of Information Eng., Electrical Eng. and Applied Mathematics (DIEM), University of Salerno, Via Ponte don Melillo, 1, 84084 Fisciano (SA), Italy

ARTICLE INFO

Available online 15 January 2014

Keywords:
Graph clustering
Graph embeddings
Graph kernels
Graph learning
Graph matching
Structural Pattern Recognition

ABSTRACT

This paper is a historical overview of graph-based methodologies in Pattern Recognition in the last 40 years; history is interpreted with the aim of recognizing the rationale inspiring the papers published in these years, so as to roughly classify them. Despite the extent of scientific production in this field, it is possible to identify three historical periods, each having its own connotation common to most of the corresponding papers, which are called here as the pure, the impure and extreme periods.

1. Motivations of the trip

The use of graphs in Pattern Recognition (PR) dates back to the early 70s; they have been used as a powerful tool for representing and classifying visual patterns, especially in structural methods, whose rationale is a vision of the objects as made of parts suitably connected to each other. Under this assumption, nodes of the graphs, enriched with properly defined attributes, can be used as descriptors of the composing parts of the objects, while the edges of the graphs represent the relationships between the parts. Good surveys of graph based techniques have been published up to now on different areas: graph-based representations, graph matching, graph edit distance, graph embedding and graph kernels [1–8] and more recently [9] provide an extensive overview of the literature over the last 40 years by introducing a detailed categorization of graph-based methods. Reading them, it is possible to have a rather exhaustive view of the scientific achievements in the above mentioned areas; the present paper, although it is a survey, has a different aim. It attempts to interpret the history of graphs by considering how, why and when they have been used in PR. It may be useful to readers, experts in PR, who are interested to better understanding the research trend on graphs when used for classification and learning of structural descriptions, instead of having a deep insight to specific methods. Surveys of applications using graphs at different levels are [10,11]; they may help the interested reader to enrich his knowledge of the field by looking how graphs can be profitably used in a wide variety of applicative fields.

The present paper surveys the literature of the graphs from the beginning to now, a history of 40 years; of course the time span of the analysis does not allow it to be exhaustive, as thousands of conference and journal papers have been produced in this period. The historical reconstruction is aimed to find out the general trends of the research in these years, and to this purpose only a few papers (over their total) have been cited here, substantially the ones that have a historical importance and characterize the approach of a new scientific branch of the field. It is worth clarifying that several papers, outstanding with respect to their scientific contributions, but less important from a historical point of view, have not been included for the evident lack of space. We are certain that their authors will understand this choice.

The use of a graph-based pattern representation induces the need to formulate the main required operations of a recognition system in terms of graphs: classification, intended as the comparison between an object and a set of prototypes, and learning, which is the process for obtaining a model of a class starting from a set of known samples, are among the key issues that must be addressed using graph-based techniques.

This paper, looking at how graphs have been used in PR along these years, assumes that it is reasonable to divide their history in three main periods (pure, impure and extreme). Of course, going into deeper details of the analysis, it would be possible to recognize further currents, and subperiods, but clear and rather objective historical trends can be highlighted by considering these three periods, as suggested also in [5], although with a different terminology.

Pure period: characterized by the fact that classification and learning problems are directly faced in the graph space, i.e. working on the graphs describing the objects at hand. The objects are classified by comparing the corresponding graphs, using suited matching algorithms. Similarly the learning is approached by building, either manually or automatically as done in [12,13], the prototype graphs, considered as a generalization of a set of graphs associated to objects belonging to the same class.

Impure period: populated by methods which transpose the basic operations defined in vector spaces on graphs, making possible the

E-mail address: mvento@unisa.it

0031-3203/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2014.01.002

reuse of effective learning and classification procedures available in the wide literature on Statistical Pattern Recognition (SPR): examples are given by the extension to graphs of Nearest Neighbor (*NN*), *K*-*NN* and $(K-K')-NN$ classifiers, Learning Vector Quantization (*LVQ*) and *K*-means clustering.

Extreme period: probably the one we are living now that is inspired by the transformation of graphs into vectors, so as to apply any present and future learning and classification procedures defined in vector spaces, as the modern kernel machines: Support Vector Machines (*SVM*), Kernel Principal Component Analysis (*KPCA*) and the Multilayer Perceptrons (*MLP*).

Starting from the consideration that sharp boundaries among these periods cannot be reliably detected, it is worth pointing out that some trends can be surely recognized. The attempt to bridge the gap between structural and statistical PR can be reliably recognized as the major driving force behind the changes across these three periods: the wide arsenal of well established methods working on vectors is too important to be neglected, and since the beginning, it solicited researchers to try its reuse on graphs. It soon appeared that an ideal approach should try to put together the advantages of both the worlds of statistical and structural PR, with the aim of exploiting their own strength and overcoming their limitations. Graphs are complex to be treated because they require high computational time and complex solutions for dealing with structure variations; at the same time vectors are significantly inadequate when the objects to be treated are organized in well identifiable structures which may simplify the above mentioned tasks. Furthermore, graphs embed smart and effective representation properties while vectors benefit from a clear mathematical support given by vector spaces.

We have just bought the tickets and are ready to start a little travel through time in the history of graphs for PR; we are going to stop at these three periods with the aim of better understanding the scientific achievements of this charming field.

2. The birth of graphs in PR

It is well known that Statistical Pattern Recognition (SPR) represents real world objects by means of a set of measures (called features): once these features (say n) have been extracted, an object becomes an n-dimensional point in the corresponding vector space. The rationale of SPR lies in the fact that the mathematical properties of vector spaces are used to face the problem of learning and classification of patterns by smart and well understandable algorithms. In fact, when the used features have been adequately selected, (for instance by discriminant analysis), an important property is supposed to hold: two points, close to each other in the vector space, correspond to similar objects in the real world; at the same time two similar objects are projected to close points in the considered vector space. It is worth highlighting the practical impact of this basic property: the vector space includes the Euclidean distance as a key tool for measuring the distance between points and it can be used as a similarity of the objects corresponding to these points. Thus, the problem of evaluating how (dis)-similar two objects are is simply brought back to evaluating the distance between the corresponding feature vectors. Fig. 1 explains this issue.

This impacting property gave impulse to the use of vector spaces in PR and plenty of statistical learning and classification algorithms have been consequently proposed since the 60′s [14]. Many researchers, due to the relatively simple underlying math, started developing and using these techniques for facing PR tasks.

Nevertheless, at the end of the 70′s an important question was raised: “Are vectors really adequate to deal with every kind of PR problem?” The answer is strictly related to the fact that vectors of a

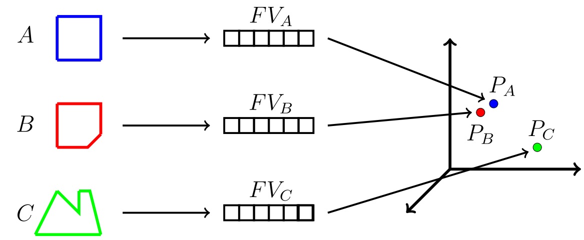

Fig. 1. Objects representation in SPR; the feature vectors are extracted and hence the objects become points of a vector space whose axes correspond to the used features.

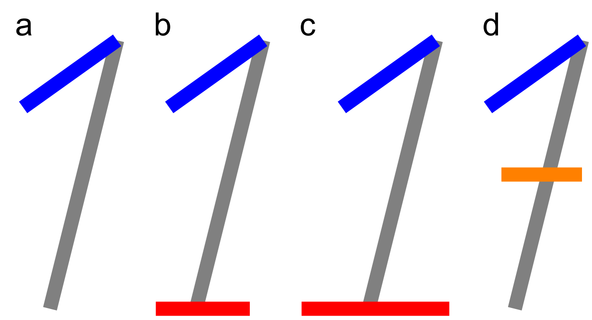

Fig. 2. The relevance of the context. (a) A sample of ‘one’. (b) The addition of a horizontal stroke at its bottom does not change the interpretation, even if it is big, as in c. (d) A little stroke added in the middle of the vertical one transforms it in a ‘seven’.

prefixed length (necessary to deal with a vector space) are not suited to represent complex patterns in which a structure made of identifiable subparts can be recognized. In fact vectors cannot adapt their length to the specific structural pattern representation; indeed the structure is expected to have a description whose length increases as the structural complexity of the input object does.

At that historical point graphs appeared to be a data structure much more adequate to represent structural descriptions of patterns: the real objects are decomposed in parts, each described in terms of a given set of parameters, and relations between these parts. The nodes and the edges of the graphs are so associated to parts and their relationships. Syntactic and Structural PR (SSPR) bet on this data structure and a great enthusiasm accompanied this novel approach [15].

In addition to their representational elegance, graphs are very suited for exploiting the contextual semantic during the classification phase: the relevance of a given feature is often related to its position inside the object and this suggests, during the comparison, to assign to each feature a weight dependent on its context, i.e. the place in which it is located. Suppose, for instance, to have the character shown in Fig. 2a: everybody interprets it as a ‘one’. If we add to this character a further stroke located at its bottom the character continues to be interpreted as a ‘one’ (see Fig. 2b: the stroke can be very big, as in Fig. 2c, but, due to its position, it is irrelevant as the interpretation of the character does not change. On the contrary, if we add even a very small stroke approximately at the center of the vertical stroke, as in Fig. 2d, it assumes a high relevance, as it induces a change of interpretation: the character becomes a ‘seven’.

Starting from this consideration it is immediate to realize the importance of the use of contextual information during the recognition phase. The graphs, differently from vector based representations, can potentially deal with this aspect, as the comparison can be somehow carried out by successive coupling of node pairs. This point will be better clarified in the rest of the paper and is crucial for catching the differences between the worlds of statistical and structural PR.