Vector space embedding of undirected graphs with fixed-cardinality vertex sequences for classification

Jonas Richiardi, Dimitri Van De Ville
Medical Image Processing Laboratory
[1]*Ecole Polytechnique Fédérale de Lausanne*
[2]*University of Geneva*
{*jonas.richiardi,dimitri.vandeville*}*@epfl.ch*

Kaspar Riesen, Horst Bunke
Institute of Computer Science and Applied Mathematics
University of Bern
{*riesen,bunke*}*@iam.unibe.ch*

Abstract

Simple weighted undirected graphs with a fixed number of vertices and fixed vertex orderings can be used to represent data and patterns in a wide variety of scientific and engineering domains. Classification of such graphs by existing graph matching methods perform rather poorly because they do not exploit their specificity. As an alternative, methods relying on vector-space embedding hold promising potential. We propose two such techniques that can be deployed as a front-end for any pattern recognition classifiers: one has low computational cost but generates high-dimensional spaces, while the other is more computationally demanding but can yield relatively low-dimensional vector space representations. We show experimental results on an fMRI brain state decoding task and discuss the shortfalls of graph edit distance for the type of graph under consideration.

1 Introduction

Graph representations offer a rich and expressive framework for data modeling and manipulation. In pattern recognition, their main advantage lies in the possibility of encoding structural information, as well as to vary object representation complexity according to need (e.g.,adding intermediate vertices in a graph depending on camera view angle). Even when structural constraints limit their degrees of freedom, graphs are a natural fit for representation and analysis of a large variety of real-world phenomena. Specifically, we concentrate on classification of a family of undirected weighted graphs having a fixed number of vertices and a fixed vertex ordering, which we call graphs with fixed-cardinality vertex sequences. This class of graphs can potentially be used for problems with multiple signals originating from a fixed number of sensors in a fixed position, such as multi-channel EEG analysis, where the number of scalp electrodes and their position over the scalp are fixed, multi-electrode array (MEA) recordings of neuron population activity, or connectivity analysis of functional magnetic resonance imaging. This latter application will be the focus of our experimental results.

In contrast to the case of vectors in normed vector spaces, where vector norms have a clear and intuitive meaning, there is no "standard" method of measuring distances between graphs. In fact, a large number of graph matching methods have been proposed to measure dissimilarity between graphs [3]. Out of these, a commonly used one, Graph edit distance (GED), is a robust and efficient measure which is defined as the sum of costs of elementary operations to transform efficiently one graph into the other. However, for the class of graphs considered here, which is subsumed by the class of graphs with unique node labels [4], no node insertion or deletion is possible, and the GED between two graphs g_1 and g_2 reduces to $d(g_1, g_2) = \big[|C_1| + |C_2| - 2|C_0|\big] + |C_0'|$, where the term between brackets accounts for the number of connections present exclusively in either of the graphs, and the C_0' set contains connections with a different label in the two graphs. We note that edge weights are not taken into account, only the count of the number of different edge weights. In the case of complete graphs, where the only operation permitted is edge substitution, GED further simplifies to $d(g_1, g_2) = |C_0'|$. In many practical situations, edge weights are noisy and thus will almost never

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.227

match, resulting in $d(g_1, g_2) \to \binom{|E|}{2}$. Thus, while GED and related similarity measures allow to compute distances between graphs in general, they are not sufficient for the class of graphs that we consider and to provide a useful metric for subsequent pattern recognition.

Embedding graphs into a vector space enables access to the rich repository of algorithmic tools from pattern analysis. A prominent class of graph embedding is based on spectral methods (e.g. [6, 2]). While spectral methods are sensitive to structural errors for general graph matching, no such problem exist for the class of graphs of interest in this paper. However, eigenspace decomposition is typically performed on the covariance matrix of training vectors made from adjacency or Laplacian matrices [10]. In domains where very little training data is available, the covariance matrix is susceptible to being very poorly estimated, yielding large reconstruction error and poor decomposition.

In this paper, we propose two vector-space embedding techniques tailored to graphs with fixed-cardinality vertex sequences; these methods can be used as a front-end to any pattern classifier. The first one, in Section 2, has very low computational complexity but generates high-dimensional vector spaces. The second one, in Section 3, is more computationally demanding but leads to lower dimensionality. We propose experimental results on a "brain decoding" task in Section 4 and conclude the paper in Section 5.

2 Direct connection label sequence embedding

A labeled simple graph $g = (V, E, \alpha, \beta)$ is a 4-tuple consisting of a set of vertices V, a set of edges E, and labeling functions α and β assigning respectively vertex and edge labels. If unique node labels exist [4], which for the purposes of graph matching is equivalent to considering the set of vertices is ordered (it is a sequence) with a fixed ordering, the "label representation" of a graph can be used. Essentially, the set of vertices V is replaced by the set of labels of vertices $L = \{\alpha(v)|v \in V\}$, the set of edges E is replaced by the set of connection between labels $C = \{(\alpha(v_i), \alpha(v_j))|(v_i, v_j) \in E\} = \{c_{ij}\}$, and the set of edge labels obtained by application of the labelling function $\beta(v)$ is directly mapped through to the sequence of connection labels Λ. The edge labelling function returns 0 if no corresponding edge is part of the edge set. Edge weights in weighted graphs can be represented directly in the adjacency matrix of the graph $\mathbf{A}$, thus making it real-valued instead of binary, or as the set of edge labels Λ. The graph g is uniquely determined by the adjacency matrix $\mathbf{A}$, and since the ordering of vertices in g is fixed and consistent because of the properties of the class of graph under consideration, $\mathbf{A}$ is also uniquely determined by g.

We propose to use the following simple embedding procedure, related to the long-vector representation of weighted adjacency matrices [5]: since $\mathbf{A}$ is symmetric, it is fully characterised by the upper triangular part. For each graph, we thus generate a vector-space embedding $\mathbf{F} : \binom{|E|}{2} \times 1$ by vectorising the edge weights of all the edges in the upper triangular part of $\mathbf{A}$. If $\mathbf{A}$ is binary and the weights are encoded as connection labels, this embedding is a bijective function $\Lambda \to \mathbb{R}^{\binom{|E|}{2}}$. Note that this differs from the embedding of [5] by not computing a covariance matrix and subsequent eigendecomposition.

The goal of this embedding is to yield a lossless vector-space representation of the graph. As the final aim is classification, dimensionality reduction and overfitting prevention is left to subsequent feature selection or to the classifier's intrinsic regularisation. In the case of non-complete graphs, vector elements corresponding to missing edges map to zero. Classifiers thus have access to the complete graph connection information.

3 Dissimilarity-based embedding

Recently, a new class of graph embedding procedures has been proposed which can be applied to both directed and undirected graphs, as well as to graphs with arbitrary labels on their vertices and/or edges, and is robust to structural errors [9]. The idea of this approach stems from the seminal work done by Duin and Pekalska [7] who proposed dissimilarities for pattern representation, recently generalized to the domain of graphs [9].

The key idea of this approach is to use the distances of an input graph g to a number of training graphs, termed *prototype graphs*, as a vectorial description of g. Assume we have a set of sample graphs, $\mathcal{T} = \{g, \ldots, g_N\}$ from some graph domain $\mathcal{G}$ and an arbitrary graph dissimilarity measure $d : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$. After selecting a set of prototypical graphs $\mathcal{P} \subseteq \mathcal{T}$, we compute the dissimilarity of a given input graph g to each prototype graph $p_i \in \mathcal{P}$. Given n prototypes, i.e. $\mathcal{P} = \{p_1, \ldots, p_n\}$, this procedure leads to n dissimilarities, $d_1 = d(g, p_1), \ldots, d_n = d(g, p_n)$, which can be

arranged in an n-dimensional vector $(d_1, \ldots, d_n)$. Thus, the mapping $\varphi_n^{\mathcal{P}} : \mathcal{G} \to \mathbb{R}^n$ is defined as the function

$$\varphi_n^{\mathcal{P}}(g) = (d(g, p_1), \ldots, d(g, p_n)), \qquad (1)$$

where $d(g, p_i)$ is any graph dissimilarity measure between graph g and the i-th prototype graph.

3.1 Dissimilarity of graphs with fixed-cardinality vertex sequences

For the class of graphs that we consider, the only differences between graphs are edge substitutions, corresponding to edge weights (connection labels) differences, and (potentially) edge deletions. We therefore define the dissimilarity measure between connection c_{ij} in the connection set C of graph g and the corresponding connection c'_{ij} in the connection set C' of graph g' as

$$d(c_{ij}, c'_{ij}) = \begin{cases} |\beta(i,j) - \beta'(i,j)| & c_{ij} \in C, c'_{ij} \in C' \\ K & otherwise, \end{cases} \qquad (2)$$

where K is a relatively large constant. This measure penalises missing connections more than connections with different edge weights.

The dissimilarity measure between a graph g with connection set $C_g = \{c_{ij}\}$ and a prototype graph p with connection set $C_p = \{c'_{ij}\}$ can then be defined as

$$d(g, p) = \sum_{i=1}^{|L|} \sum_{j=i+1}^{|L|} d(c_{ij}, c'_{ij}) \qquad (3)$$

3.2 Prototype selection

Provided sufficient training data is at hand, the selection of the n prototypes $\mathcal{P} = \{p_1, \ldots, p_n\}$ is a critical issue in graph embedding since not only the prototypes $p_i \in \mathcal{P}$ themselves but also their number n affect the resulting graph mapping $\varphi_n^{\mathcal{P}}(\cdot)$ and thus the performance of the corresponding pattern recognition algorithm, as reported in various papers [9, 8]. Because of the dearth of data in our application, we currently use all training graphs (in cross-validation) as prototypes. Feature selection can be carried out in vector space as a later stage to eliminate uninformative or noisy components of the dissimilarity vector [8].

4 Classification experiments

The dataset consists of spatiotemporal fMRI data from 15 subjects (GR-EPI sequence, TR/TE/FA = 1.1s/27ms/90°, matrix = 64×64, voxel size = 3.75×3.75×4.2mm^3). Stimulation was done with alternating blocks of movie excerpts (50s) and resting periods (90s). There are 9 movie blocks and 9 resting blocks. We concatenate blocks from the same condition after linear detrending, and use whole-brain atlasing (90 regions) to average timecourses, yielding 90 time-series/subject/condition. Each time-series is filtered into four wavelet subbands using a non-subsampled orthogonal discrete wavelet transform (4 vanishing moments). Pairwise temporal correlation (Pearson product-moment) is computed on each of the filtered time-courses, yielding in total four 90×90 correlation matrices reflecting co-activation of spatially distinct brain regions at different time scales.

The 90 regions can be represented as vertices of a functional connectivity graph, while the correlation coefficients correspond directly to edge weights or edge labels. Thus, for each subject, 4 subband-specific graphs are obtained for each condition (*resting* and *movies*). The two-class classification problem is to be able to infer whether the subject is watching a movie or resting based on the connectivity graphs only.

For both embeddings, we perform tests with two types of classifiers: an ensemble of 21 boosted functional trees classifiers (FT21), and a linear kernel SVM using within-fold cost parameter optimisation (SVM). For dissimilarity-based embedding, we perform additional tests using a multi-layer perceptron (MLP) trained with backpropagation. All classifiers are implemented in Weka [11]. Stratified leave-one-subject-out accuracy results are given in Table 1, where 100% accuracy is for 30 graphs, 15 from each class.

Graphs constructed in lower-frequency subbands are presumably easier to classify because resting-state connectivity is known to be particularly strong at these frequencies. The higher accuracy obtained overall by direct connection label sequence embedding can be attributed to the very small amount of training data, yielding a low number of prototype graphs to use for dissimilarity-based embedding (DBE). Considering the low dimensionality of DBE (28), performance is quite remarkable.

5 Conclusion

In this paper, we have shown that when graphs have fixed-cardinality vertex sequences, existing general-purpose graph-theoretical methods

Subband	Classifier	DE	DBE
1 (0.23-0.45 Hz)	SVM	53%	53%
	FT21	53%	57%
	MLP	—	53%
2 (0.11-0.23 Hz)	SVM	87%	60%
	FT21	80%	60%
	MLP	—	63%
3 (0.06-0.11 Hz)	SVM	93%	83%
	FT21	93%	77%
	MLP	—	87%
4 (0.03-0.06 Hz)	SVM	97%	83%
	FT21	83%	67%
	MLP	—	83%

Table 1. Accuracy for different embeddings. DE: direct connection label sequence embedding, DBE: dissimilarity-based embedding.

for graph matching, such as GED, may theoretically fail. We suspect that this is also the case for other commonly used methods such as maximum common subgraph (MCS); in this particular case there are conditions under which MCS is equivalent to GED [1], so it is unlikely that graphs in this family will be successfully classified. We have proposed two vector-space embeddings that allow effective classification of this type of graphs. The first one is lossless (but high-dimensional) and requires subsequent feature selection or regularisation of the classifier. Given a prototype set smaller than the number of vertices in a complete graph, the second one intrinsically reduces the dimensionality using the recent concept of dissimilarity-based representation. We have demonstrated the feasibility of both approaches using experimental results on an fMRI brain decoding task. In the field of brain decoding, the use of connectivity measures for classification is novel with respect to existing classification techniques that rely on voxels' activation levels. It is likely that this kind of approach will be useful for cognitive neurosciences (insight in interactions between brain regions) and for clinical practice (alterations in functional connectivity are present in a variety of brain disorders such as Alzheimer's disease, schizophrenia, and others).

Future work will include testing the approach on other datasets such as whole-scalp EEG recordings and geoscience datasets, exploring new embeddings for this type of graphs, and empirical and theoretical comparison with other embedding techniques such as spectral methods.

Acknowledgements

We thank H. Eryilmaz, S. Schwartz, and P. Vuilleumier for providing the fMRI data. This work was supported in part by the Swiss National Science Foundation (grants PP00P2-123438 and 200021-113198/1), in part by the Société Académique de Genève and the FOREMANE foundation, and in part by the Center for Biomedical Imaging (CIBM) of the Geneva and Lausanne Universities, EPFL, and the Leenaards and Louis-Jeantet foundations.

References

[1] H. Bunke. On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters*, 18(8):689–694, Aug. 1997.

[2] T. Caelli and S. Kosinov. Inexact graph matching using eigen-subspace projection clustering. 18(3):329–355, 2004.

[3] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(3):265–298, 2004.

[4] P. Dickinson, H. Bunke, A. Dadej, and M. Kraetzl. Matching graphs with unique node labels. *Pattern Analysis & Applications*, 7(3):243–254, Sept. 2004.

[5] B. Luo, R. Wilson, and E. Hancock. Eigenspaces for graphs. In *Proc. 5th Asian Conf. on Computer Vision (ACCV)*, pages 487–492, Melbourne, Australia, January 2002.

[6] B. Luo, R. Wilson, and E. Hancock. Spectral embedding of graphs. 36(10):2213–2223, 2003.

[7] E. Pekalska and R. Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*. World Scientific, 2005.

[8] K. Riesen and H. Bunke. Non-linear transformations of vector space embedded graphs. In A. Juan-Ciscar and G. Sanchez-Albaladejo, editors, *Pattern Recognition in Information Systems*, pages 173–186, 2008.

[9] K. Riesen and H. Bunke. Graph classification based on vector space embedding. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 23(6):1053–1081, 2009.

[10] R. Wilson, E. Hancock, and B. Luo. Pattern vectors from algebraic graph theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(7):1112–1124, 2005.

[11] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufman, 2nd edition, 2005.