Graph Kernels and Support Vector Machines for Pattern Recognition

Léo Andéol[1]
Supervised by: Prof. Hichem Sahbi

Master DAC - Sorbonne Université

May 2019

[1] leo.andeol@gmail.com

Summary

Motivation

- A lot of data can be represented as graphs such as proteins
- Being able to compare them would be useful (classification, clustering)

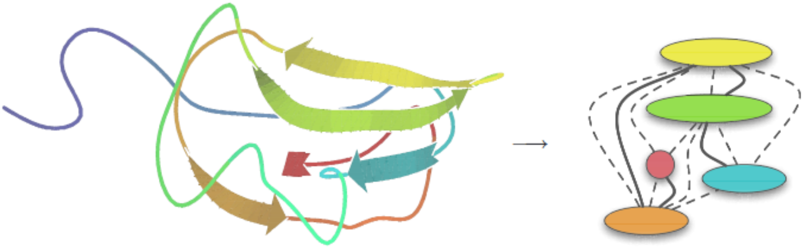


Figure 1: A fragment of a protein transformed into a graph[7]

Current Methods

Support Vector Machine

SVMs are models used in classification introduced about 25 years ago. They have several advantages

- Great accuracy
- Great capacity of generalization
- Allows the use of kernels in its dual form

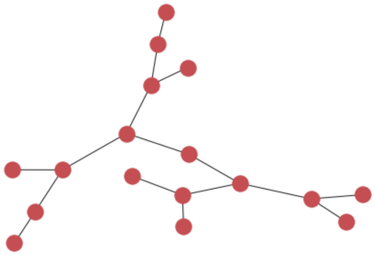
Kernels

The kernel trick can replace the dot product while implicitly projecting data to a feature space and combine very well with the SVMs

- Computes data projection faster implicitly (ex. RBF kernel)
- Improve the accuracy of SVM by making linear separation easier

Objective

- These methods are adapted to **vector data**
- Graphs and their adjacency matrices aren't and vectorizing implies a loss of information
- A **new** type of kernel is required



$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Figure 2: A tree graph and an adjacency matrix

Summary

Background : graphs

Definition

A graph[1] is a type of mathematical structure that represents connections between objects. It is more precisely an ordered pair $G = (V, E)$ of two sets: vertices V (or nodes) and edges E that connect two vertices together.

$$E \subseteq \{(u, v) : (u, v) \in V^2\}$$

Properties

- Undirected
- Labeled or not
- Degree
- Path and Cycle
- Connected
- Tree
- Subgraph
- Dual Graph

Background : support vector machines

SVM

Background : kernels

Definition

In its dual form, the SVM problem only requires a dot product between the observations' vectors.

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}$$

This means the vectors can be mapped to higher dimensions with a function ϕ. Moreover, even the dot product itself can be replaced by a function κ without explicitly specifying the map ϕ as long as the function is positive semi definite.

$$\kappa(\mathbf{x_i}, \mathbf{x_j}) = e^{-\frac{\|\mathbf{x_i}-\mathbf{x_j}\|^2}{2\sigma^2}}$$

An example of kernel : the RBF kernel

Graph Kernels

Graphlets

Random walks

Definition

kernel def

Acceleration methods

Cas particulier

Inverse Kernel

inv ker

On va vouloir accelerer machin ce kernel va être l'objectif de plusieurs méthodes

Sylvester Equation

content...

Conjugate Gradient

content...

Fixed Point

content...

Spectral Decomposition

content...

Nearest Kronecker Product Approximation

- The idea is to approximate two matrices S and T such that $\|W_\times - A \otimes B\|_F$ is minimized.
- Labeled-graph kernel computation can be turned into into an unlabeled one with some loss in accuracy, but gain in computation time.
- Computed in $O(dn^2)$ time
- All methods such as Spectral Decomposition can then be applied

Summary

Frame

Conclusion

conclu

References

[1] J. A. Bondy, U. S. R. Murty, *et al.*, *Graph theory with applications*, vol. 290. Citeseer, 1976.

[2] K. M. Borgwardt, C. S. Ong, S. Schönauer, S. V. N. Vishwanathan, A. J. Smola, and H.-P. Kriegel, "Protein function prediction via graph kernels," *Bioinformatics (Oxford, England)*, vol. 21 Suppl 1, pp. i47–56, June 2005.

[3] W. Imrich and S. Klavzar, *Product graphs: structure and recognition*. Wiley, 2000.

[4] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Min. Knowl. Discov.*, vol. 2, pp. 121–167, June 1998.

[5] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley-Blackwell, Oct. 1998.

[6] Y. Nesterov, *Lectures on Convex Optimization*. Springer Optimization and Its Applications, Springer International Publishing, 2 ed., 2018.

[7] S. V. N. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt, "Graph Kernels," *Journal of Machine Learning Research*, vol. 11, no. Apr, pp. 1201–1242, 2010.

[8] N. Shervashidze, S. V. N. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt, "Efficient graphlet kernels for large graph comparison," in *Artificial Intelligence and Statistics*, pp. 488–495, Apr. 2009.

??

??