THESIS @ENSICAEN (FRANCE)

ENSI CAEN
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

The **GREYC Image** of the ENSICAEN is among the biggest research groups at the ENSICAEN, specialised in the definition of novel methodologies for both structural and statistical pattern recognition.

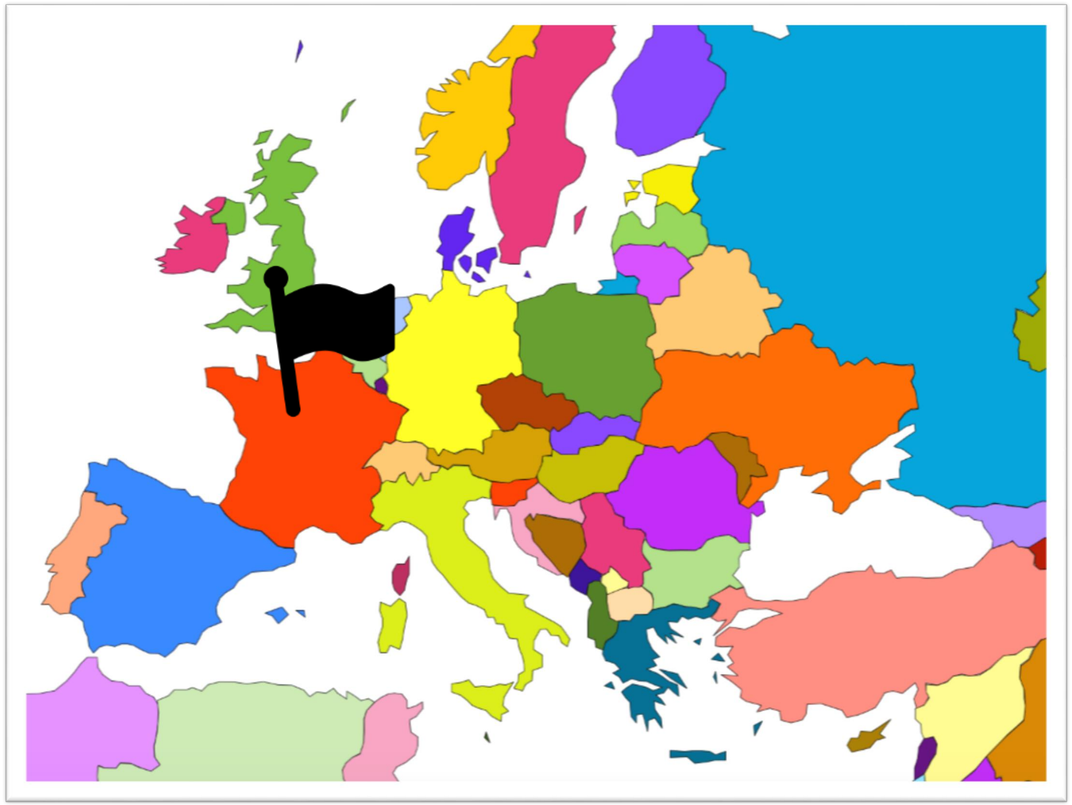
Luc Brun, the contact professor for the ERASMUS@ENSICAEN, is currently the Vice Chair of the IAPR TC15 (Technical committee in Graph-based Representations in the Pattern Recognition of the International Association for Pattern Recognition).

Prof. Brun in numbers: H-Index: 25; Citations: 2123

Required Skills for the thesis:

Machine Learning e Big Data Analytics

Give a look to the Lab Website



UNIVERSITÀ DEGLI STUDI DI SALERNO

Cosmo D’Antuono

ENSICAEN

A GRAPH BASED METHOD TO RECOGNISE SCHIZOPHRENIA IN THE HUMAN BRAIN

Abstract

Nowadays a lot of information are represented and processed in the form of graphs. Our social interactions, be they personal relationships, political associations, financial transactions, professional collaborations, the spreading of rumors or diseases, or physical transport and travel, increasingly occur within graph or networks that are evolving across time.

All of these networks are examples of complex systems, with highly structured connectivity patterns, multiscale organization, nonlinear dynamics, resilient responses to external challenges, and the capacity for self-organization that gives rise to collective or group phenomena. Modern developments in graph theory and complex systems have delivered important insights into the structure and function of these diverse networks, as well as quantitative models that can both explain and predict network phenomena. In the last decades, there has been an explosion of studies on network modeling of brain connectivity.

This thesis is focused, mainly, on recent findings concerning graph theoretical analysis of human brain networks with a variety of statistical analysis proposing a method that will lead us, in the near future, to identify reliable biomarkers for disease diagnosis and treatment of human brain. The exciting research in brain graph led to integrative and comprehensive descriptions of structural and functional organization of the human brain, which provides important implications for health and disease. The aim of the thesis is to propose a method that would allow us to distinguish, through the analysis of data related to the brains of a population of subjects, Individuals with mental illness from healthy individuals.

The first studies were carried out with regard to the extraction and selection of features. In this phase it has been made extensive use of statistical analysis methods such as the Principal Component Analysis (PCA). So tests were made on the distributions of these graphs in order to determine which was the best way to obtain a good discrimination between the two classes of graphs like Anderson-Darling test, Mann-Whitheny-Wilcoxon test and the Kruskal-Wallis test. After this phase, on the basis of the results obtained, it was carried out the training of a support vector machine (SVM), which, through the adoption of different kernels and thresholding indices, has allowed the identification of which were the best values to use to operate the classification.

Finally was conducted a classification no longer on graphs but on hierarchies using the technique of multiple kernel learning (MKL) with the aim of being able to isolate typical markers of a brain suffering from mental illness. At the end of the description of these three procedures will then be shown the results obtained and the considerations observable.

UNIVERSITÀ DEGLI STUDI DI SALERNO

Diego Maiorano

ENSICAEN

USE OF THE QUADRATIC ASSIGNMENT TO COMPUTE THE GRAPH EDIT DISTANCE OVER MOLECULAR DATASETS

Abstract

Graphs are the most suitable representation in many application fields where the data are naturally represented as entities and relationships among them, such as molecular biology, chemistry, social networks and so on. In these fields there is the need of process and analyze a huge quantity of data collected during the years. During the process of such data, a frequent problem is to search for structural similarities because it is supposed that similar structure will show similar properties, especially in chemistry.

A graph matching problem can be solved finding a correspondence between vertices and edges of two graphs and can be used for solve a lot of problem such as subgraph or graph isomorphism and maximum common subgraph. Usually graph matching methods are divided into two categories: exact matching methods, that require a strict correspondence, and inexact matching methods, where instead the structure can be partially different.

raph Edit Distance (GED) is an inexact graph matching method that defines a similarity metrics between two graphs by considering the sequence of edit operation (addition, deletion and substitution of edges and vertices) needed to transform the first graph in the second one. Unfortunately, despite this distance is very useful, the computation of the exact value is an NP-Complete problem that is usually addressed by mean of the A* algorithm. To avoid this drawback the GED is commonly approximated by using different algorithms like Beam Search, Genetic Algorithms, Convex Optimization Algorithms and so on. Actually, a very interesting method has been propose by K. Riesen and H. Bunke and is based on the formulation of the GED as a Weighted Bipartite Graph Matching Problem engaged as a Linear Sum Assignment Problem (LSAP).

It has been proven to provide interesting results in different application field, especially for the time complexity that is cubic with respect to the size of the two graphs. Despite this, the accuracy of the GED provided is strongly limited by the approximation needed to solve the problem as a LSAP. The aim of this thesis is to improve the accuracy by performing the calculation of GED as an assignment problem with a quadratic objective function, namely the Quadratic Assignment Problems (QAP). Thus, firstly we have formalized the GED as a QAP by a new cost function to encodes the overall structure of the graph in the quadratic term.

Then we have used two different algorithms to solve the QAP, i.e. the Soft Assign and the Integer Projected Fixed Point. The former is a discrete time and continuous state algorithm, proposed by Rangarajan, Yuille, Gold and Mjolsness. The latter has been proposed by Leordeanu, Hebert and Sukthankar. It finds the optimal continuous solution by ignoring, during optimization, the original discrete constraints.

Finally we have compared the results obtained by these algorithms with those obtained from other state of the art methods based on the LSAP, in particular we have referred to an analysis produced by Gauzere, Carletti, Brun and Vento. In both works, we have extracted data from some datasets of molecular structures containing labeled or unlabeled and cyclic or acyclic graphs. The experiments have confirmed the effectiveness of the QAP method that has provided a better approximation of the GED over all the datasets, with an improvement for both execution time and value of edit-distance.

.DIEM

UNIVERSITÀ DEGLI STUDI DI SALERNO

MIVIA

Daniele Santorelli

ENSICAEN

A FACE VERIFICATION METHOD BASED ON THE GRAPH EDIT DISTANCE

Abstract

Computing the similarity between graphs is a challenging problem in Pattern Recognition; it is commonly addressed by using the Graph Edit Distance (GED). The latter represents a dissimilarity measures between two graphs, defined as the cost of the cheapest sequence of edit operations (such as substitution, deletion and insertion of node or edge), aimed to transform the first graph in the second.

The exact value of GED is usually computed by using the A* algorithm whose computational complexity is NP-Hard. So that, in the last decade, several approaches have been proposed to approximate the GED. One of the most interesting, proposed by K. Riesen and H. Bunke [1], is based on approximating the GED as a Linear Sum Assignment Problem (LSAP).

The main problem of the Bunke's approach lies in the accuracy of the approximation, because this method considers only the cost related to the nodes and includes in the latter also the one of the edges. A new approach approximates the GED by using the Quadratic Assignment Problem (QAP). It computes the exact value of the GED by considering both the nodes and the edges of the graphs without any approximation in the definition of the cost function. Unfortunately, solving the QAP is an NP-Hard problem, so it is generally approximated and consequently also the GED.

In our work we have approximated the QAP by using the Integer Projected Fixed Point (IPFP), a continuous optimization algorithm originally based on the Hungarian algorithm. In order to improve its accuracy and reduce the execution time we have substituted the Hungarian with the Sinkhorn's algorithm. Indeed, the computational complexity of Sinkhorn O(n log n) is smaller than the computational complexity O(n3) of Hungarian.

Once the method to approximate the GED has been assessed, we have used it on the face verification problem, whose purpose is to authenticate a person by analyzing his face. To this aim we have defined and test different graph-based representation for the face combining the information obtained from a set of fiduciar points, a set of descriptors (LBP, SIFT a SURF) characterizing the aspect of each point and the structure obtained connecting these points. As a matter of fact, we can divided our work in two parts: firstly we have used the Sinkhorn's algorithm instead of the Hungarian algorithm in the IPFP, then in the second part of this work the IPFP has been used to realize a face verification system.

UNIVERSITÀ DEGLI STUDI DI SALERNO

Antonio Salvati

ENSICAEN

A GRAPH-BASED METHOD TO RECOGNIZE PEOPLE GENDER

Abstract

This thesis has been focused on two topics: firstly, defining an efficient method to computed the Graph Edit Distance (GED), then, realizing a graph-based algorithm to recognize the gender of a person. The GED is an inexact graph matching method used to compute the similarity between two graphs. It is defined as the cost of the cheapest sequence of edit operations (addition, deletion and substitution of edges and vertices) needed to transform one graph in the another. This distance is very useful to apply Pattern Recognition methods on graphs, but the computation of its exact value is an NP-Complete problem, usually addressed by means of the A* algorithm. Likely in many practical applications an approximation of the GED is suitable to perform the task.

So that, several methods have been proposed to compute approximation of the GED in specific contexts. A noteworthy method has been recently proposed by K. Riesen and H. Bunke. It is based on the formulation of the GED as a Weighted Bipartite Graph Matching Problem and then solved as a Linear Sum Assignment Problem (LSAP). This main limit of this formulation lies the linear cost function that does not consider directly the cost of the edges, but embeds them in the cost to map the nodes. A new formulation, based on a quadratic cost function, has been proposed in the technical report "A Quadratic Assignment formulation of the Graph Edit Distance" S. Boegleux, V. Carletti, L. Brun, B. Gauzerre, P. Foggia, M. Vento. In this method the GED is still defined as Weighted Bipartite Graph Matching Problem, but the latter is solved as Quadratic Assignment Problem (QAP), so as to consider also the edges in the quadratic term of the cost function. In this thesis we have focused our attention on a new algorithm to approximate the QAP by using a modified version of the Integer Projected Fixed Point (IPFP) algorithm. The latter uses a continuous optimization approach and solves the QAP using a in two main steps. In the first step the algorithm computes a discrete linear approximation of the quadratic cost function around the current solution by using the Hungarian algorithm to solve a LSAP; in the second step the algorithm computes the next candidate solution by solving the relaxed problem. The Hungarian algorithm finds out the optimal solution of the LSAP, but has a cubic time complexity. In order to reduce the overall computational complexity of the IPFP we have substituted the Hungarian algorithm with the Sinkhorn's algorithm. The latter is commonly used to solve transportation problems, but it is also used to approximate the LSAP and requires a less than quadratic computational complexity. Therefore, to confirm the efficiency of our algorithm, we have compared it with the original one and other state of the art algorithms used to approximate the GED.

As regards the second topic, we have used the algorithm introduced above to recognize the gender of a person whose face has been represented as a graph. In order to perform this task we have followed these steps: pre-elaboration on the training set of images, fiducial points extraction, fiducial points description (by adding additional features), generation of a graph by connecting the points.

UNIVERSITÀ DEGLI STUDI DI SALERNO

Salvatore Sessa

ENSICAEN

A NEW METHOD TO APPROXIMATE THE GRAPH EDIT DISTANCE APPLIED TO THE FACE VERIFICATION PROBLEM

Abstract

In the last years, graph-based representation has been widely used in many application fields where the data are naturally represented as entities and relationships among them, such as molecular biology, chemistry, social networks and so on. So that, interesting challenge in Pattern Recognition is defining an efficient method to measure the distance between two or more graphs. A commonly used measure is the Graph Edit Distance (GED); it represents a measure of dissimilarity between two graphs defined as the cost of the cheapest sequence of edit operation (such as insertion, deletion and substitution of edges and nodes) aimed to transform one graph into another. The computation of the exact value of the GED is expensive because its computational complexity is exponential with respect to the number of nodes; to this reason the computation of the exact GED is suitable only for small graphs. This is the reason why, in the last decade, several approaches have been proposed to approximate the GED, one the of the most interesting has been recently proposed by K. Riesen and H. Bunke, and is based on solving a Bipartite Matching as a Linear Sum Assignment Problem (LSAP). Despite this method provides good results in different application, there is a limit in its accuracy that is related to how the cost function of the LSAP is defined.

Indeed, the latter can consider only the costs to related to nodes and has to approximate the costs of the edges, that are included in those of the nodes. A more recent method proposes to substitute the linear formulation of the cost function with a quadratic one and to compute the GED by solving a Bipartite Matching as a Quadratic Assignment Problem (QAP). Thus, the cost function does not carry any approximation because the cost related to the edges is defined by using the quadratic term. Another important advantage lies into to the fact that solving optimally the QAP will provide the exact value of the GED; note that this is not true for the LSAP.

Unfortunately, the QAP is an NP-Hard problem, and its solution can only be approximated. In order to perform this approximation, we have realized an optimized version of the Graduated nonConvexity and Concavity Procedure (GNCCP). This approach is a general optimization framework to approximate combinatorial optimization problems defined on the set of partial permutation matrices. GNCCP comprises two sub-procedures, graduated non-convexity, which realizes a convex relaxation, and graduated concavity, which realizes a concave relaxation. GNCCP realizes exactly a type of convex-concave relaxation procedure (CCRP), but with a simpler formulation, without convex or concave relaxation in an explicit way. Thus, we have realized a new algorithm aimed to obtain a better approximation of the QAP (and so of the GED). The results of the obtained algorithm show an improvement in the accuracy of the approximated GED but an increase of the execution time.

In a second stage, the algorithm has been used to solve the problem of face verification, also called face authentication, that consists in deciding if a person is contained in a knowledge base (by analyzing his face) and verifying if this person is who he/she claim to be. People faces have been represented as graph using different aspect descriptors to characterized the nodes and the edges of these graphs. So, we have used the GED, computed with our algorithm, to measure the similarity between two faces. These distances have been used by a k-NN classifier to perform the verification task. In order to select the best graph-based representation we have compared them on a dataset of more than 120 faces of 12 different people

.DIEM

UNIVERSITÀ DEGLI STUDI DI SALERNO

MIVIA

Giovanni De Santis

ENSICAEN

AN ANOMALY DETECTION APPROACH TO DETECT FALLS

Abstract

Falls are defined as events which result in a person coming to rest inadvertently on the ground or floor or other low level. They represent a problem related to all ages, genders or types of activities and they can lead to serious injuries and even to death.

Given the potential dangerousness of these events, in the last years there was been a growing interest in the research and industrial communities in defining methods able to automatically detect falls and noticeably reduce the response time of the medical staff or caregivers when a fall occurs.

Over time, this increasing effort has conducted to the proposal of several distinct solutions based on different varieties of sensors (camera-based sensors, proximity sensors, wearable sensors, hybrid sensors, etc.) and different kind of algorithms (simple threshold-based algorithms and machine learning algorithms).

Nonetheless, despite all the efforts, falls cannot be detected efficiently yet. The reasons at the basis of this difficulty are variegated, but the most relevant one lies in the difference existing between the laboratory environment in which the fall detection systems are designed and the real environment in which they have to be deployed and used. Real falls, in fact, are abnormal activities that occur rarely.

Thus, in order to collect the necessary data for the conception of the generic fall detection mechanism, researchers have always resorted to simulated falls (falls artificially performed by young individuals). But the diversity existing between the design settings and the real usage environment has caused a significant difference between the recorded laboratory performance and the real-world one.

Hence, in order to face this issue and reduce its effects, a new research branch based on anomaly detection machine learning has recently started. According to this branch, falls are treated as what they really are: anomalies. As a consequence, the generic fall detection method is designed on the basis of normal Activities of Daily Living (ADL) and a fall is defined as an activity “distant” from the normal ones.

Such an approach has appeared to us as the most promising research direction along which proceed. Our purpose, in particular, has been to analyze the application of a selected group of anomaly detection algorithms to the general problem of fall detection and extend their utilization to more dangerous and complicated situations occurring as a person is practicing a sport.

.DIEM

UNIVERSITÀ DEGLI STUDI DI SALERNO

MIVIA

Roberto Pacilio

ENSICAEN

FEASIBILITY STUDY OF A MULTI-CAMERA VIDEO ANALYSIS SYSTEM FOR "MONDEVILLE 2 SHOPPING CENTRE" OF CAEN

Abstract

Fare impresa oggi non significa più affidarsi solo ed unicamente all'eccellenza individuale o all'intuito imprenditoriale che riesce a sfornare idee valide per sfondare nel sempre più vasto mercato delle iniziative di business. Il marketing non è più frutto della creatività intellettuale ma si fonda sempre più sulla capacità di intraprendere iniziative analitiche. In questi ultimi venti anni, il rispondere a questa capacità/necessità oggettiva di osservare come si muove il business di un'azienda è diventato il ruolo essenziale dei moderni sistemi di Business Intelligence. Il ruolo della Business Intelligence non è soltanto quello di monitorare come si comportano i propri clienti, come reagiscono alle offerte di mercato, come rendere più appetibili le proposte di prodotti/servizi, ma anche monitorare e quindi migliorare l'esecuzione dei processi produttivi.

Questo lavoro si pone come obiettivo quello di produrre uno studio di fattibilità per un sistema di Business Intelligence che possa supportare il management del centro commerciale Mondeville 2 – Centre Commercial de Caen (Francia) nel prendere decisioni strategiche. Inoltre, un prototipo del sistema di Business Intelligence proposto è stato sviluppato affinché il committente ne possa apprezzare i risultati. L'output finale del sistema è caratterizzato da due indicatori: (1) punti caldi e punti freddi identificati all'interno di un'area del centro commerciale; (2) percorsi tipici dei clienti all'interno di detta area.

.DIEM

UNIVERSITÀ DEGLI STUDI DI SALERNO

MIVIA

Vincenzo De Notaris

ENSICAEN

A MULTI-CAMERA TRACKING ALGORITHM DEVISED FOR A BUSINESS INTELLIGENCE APPLICATION IN A SHOPPING CENTRE BY VISUAL DATA

Abstract

This work has been realised starting from a request by the Mondeville 2 – Centre Commercial de Caen3, interested in having a Business Intelligence system based on Computer Vision techniques to inspect walking paths of the customers inside an area of interest.

To inspect walking paths by visual data, the BI system have to perform the following operations:

1. detect and track people inside the field of view for each camera. For each person, we have to assign an unique label and store his walk;
2. merge all trajectories computed by each single cameras to obtain global trajectories, guaranteeing the labelling correctness for moving persons inside the entire area of interest;
3. process global trajectories to produce data which characterise typical walking paths, useful to define business strategy for the shopping centre.