

Séminaires doctorants 4

14 et 15 juin 2006

Sessions Master 2 Recherche

Actes édités par l'association des doctorants du campus STIC.

Les travaux individuels publiés restent l'unique propriété de leurs auteurs. La copie et la distribution de ces actes dans leur intégralité, cette notice comprise, sont toutes deux autorisées.

Table of Contents

Formal Verification of Exact Real Arithmetic and Analytic Functions Nicolas Julien	1
Semi-supervised and Unsupervised Classification: Applications of the GAIA Data	2
Face Recognition with Video Usman Saeed	4
On Symmetric Sandpiles <i>Theophilos Pisokas</i>	5
Error Mining in Parsing Lionel Nicolas	6
State Based Evolutionary Algorithm Sébastien Tesquet	7
A Composition Mechanism Applied to Business Models	8
Contextual Spatial Ontology for Interoperability of Geographic Web Services $Florian\ Prud\ homme$	9
Identifiability and Separability by Particle Filtering Benoît Lagadec	10

Formal Verification of Exact Real Arithmetic and Analytic Functions

Nicolas Julien

INRIA Sophia Antipolis, France julien@polytech.unice.fr

The aim of this work is to verify formally the mathematical proofs of correctness for a library of exact real arithmetic operations. A real number is represented as an infinite sequence of signed digits. The proof assistant Coq allow us to use such a representation and prove the functions we define thanks to co-inductive types and co-recursive functions and proofs.

Semi-supervised and Unsupervised Classification: Applications of the GAIA Data

Laurent Galluccio *

Laboratoire Universitaire d'Astrophysique de Nice, France laurent.gallucio@unice.fr

Abstract. During previous work, different approaches have been tried to solve the problem of unsupervised classification, where we have no prior knowledge on the classes we want to obtain. The aim of my training period is to determine the hardness of some classes of solutions that build a taxonomic classification on asteroids from astrophysics datasets. The methods we develop during this period are: segmentation of a graph based on the Minimal Spanning Tree (Prim's algorithm) and spectral clustering methods. We have tested our methods on different surveys to validate them so we can simulate data from Gaia, knowing the filters bands of Gaia.

1 Introduction

The mission Gaia (Global Space Astrometry mission) will be launched in 2011 by the ESA from Kourou (Guyana). The aim of this mission is to establish a 3D map of our solar system. The discovery of one billion stars is hoped for.

A secondary mission will be the detection and the measurements of one to ten millions of asteroids (only 250 000 known nowadays). Incorporating a classifier to make clusters of asteroids inside the satellite is indispensable.

2 Similarity Between Specters

The data we have on our disposal are the reflectances of the asteroids. These objects are reflecting objects. Thus, we measure the percentage of incident light of the sun reflected by the asteroid in order to determine the surface's mineralogical composition. Asteroids belonging to a same class will have similar specters on reflectance.

We have to characterize the similarity of these specters. We define an affinity matrix W; or on the same way a matrix of distance D (dimension $N \times N$, if we have N specters). The matrix W can be defined with D (most of the time: $w_{ij} = \exp(-\alpha d_{ij})$), if the distance between two specters is low, similarity is high.

^{*} joint work with Olivier Michel & Philippe Bendjoya

3

We use different metrics, the most used one being the Euclidean distance: $d(A_i, A_j) = \sqrt{\sum_{k=1}^{P} (x_{ik} - x_{jk})}$. We can also use some informational divergence. We have just considered Csizr divergence, which is defined as

$$h\left[E_{p_j}\left\{g(\frac{p_{ik}}{p_{jk}})\right\}\right].$$
(1)

We use essentially Renyi and Kullback Leibler's divergences.

3 Methods for Clustering the Dataset

At first, we build the Minimal Spanning Tree with the Prim's algorithm and an improvement of the KMeans algorithm (also known as Lloyd algorithm) [1, 2]. Considering the distance between vertices on the built tree, we form clusters that contain points with high similarity. Some points with low similarity with the existing clusters are not indexed in any cluster. The use of the KMeans algorithm, with the knowledge brought by the MST, allows to add these points to one cluster.

The other method used during my training period is spectral clustering [3], which relies on the eigenstructure of a similarity matrix W to partition points into R disjoint clusters. The quantity we want to minimize is the NCUT as defined in [3]. Jordan et Bach [4] have shown that minimizing this function is equivalent to minimizing a weighted distortion measure. They used a weighted form of KMeans.

References

- Hero, A.O., I., Ma, B., Michel, O., Gorman, J.: Applications of entropic spanning graphs. IEEE Signal Processing Magazine 19 (2002) 85–95
- Michel, O.J.J., B.P., Rojoguer, P.: Unsupervised clustering with mst: Application to asteroid data. (In: PSIP 2005) Paper 01_03.
- 3. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 888–905
- Bach, F.R., Jordan, M.I.: Learning spectral clustering. Technical report, EECS Department, University of California, Berkeley (2003)

Face Recognition with Video

Usman Saeed

Institut Eurécom, Sophia Antipolis, France saeed@polytech.unice.fr

Still images have been the dominant modality for research in face recognition, but with the rapid increase in the use of video surveillance equipment and webcams, recognizing people using video sequences has started to attract the attention of the research community. Most of these still image systems use only appearance for automatic face recognition, and the behavioral aspect has been completely ignored that can be used for discriminating identities. Then, most of these strategies have been developed using perfectly normalized image databases, but for actual applications it would be better to work on real data; for example, low quality compressed sequences or video surveillance shots.

In this internship, we propose a new person recognition system based on temporal signals of features from rigid global and non rigid local face dynamics. The local face dynamics consist of features relating to non rigid eyes and mouth motion. Rigid face movement is analyzed by calculating the degree of face symmetry and angle of the face in each video frame. Statistical features are then computed from these signals, in order to characterize the motion information from the video, and used for discriminating identities; the classification task is done using a Gaussian Mixture Model (GMM) approximation and Bayesian classifier.

On Symmetric Sandpiles

Theophilos Pisokas

Laboratoire I3S, Université de Nice - Sophia Antipolis, France pisokas@polytech.unice.fr

Self-Organized Criticality (SOC) is a very common phenomenon that can be observed in Nature. It concerns, for example, sandpiles formation, snow avalanches, stock market crashes and so on.

A symmetric version of the well-known SPM model for sandpiles is introduced. We prove that the new model has fixed point dynamics. Although there might be several fixed points, a precise description of the fixed points is given. Moreover, we provide a simple closed formula for counting the number of fixed points originated by initial conditions made of a single column of grains. Finally, we study the time complexity of the system evolution.

Error Mining in Parsing

Lionel Nicolas

Laboratoire I3S, Université de Nice - Sophia Antipolis, France INRIA Rocquencourt, France nicolasl17@yahoo.fr

Improving the syntactic coverage of a natural language grammar and the quality of its lexicon are difficult tasks because of the richness of a human language like French. For this reason, we explore feedback mechanisms that extract information from the parsing of large corpora (for instance, around 400 000 sentences from a journalistic corpus). The basic idea is to compute the parsing error rate for a word, a lemma or a sequence (of lemma or syntactic categories) in order to identify those that are significantly below the expected average error rate. Words with high error rates tend to indicate incorrect or incomplete information in the lexicon while sequences with high error rates tend to indicate problems with the grammar.

The main objective of this work is to explore methods for suggesting potential corrections for words with high error rates in order to complete or correct a lexicon.

State Based Evolutionary Algorithm

Sébastien Tesquet

Laboratoire I3S, Université de Nice - Sophia Antipolis, France tesquet@polytech.unice.fr

We will present to you a new model of Evolutionary Algorithm: State based Evolutionary Algorithm. The Evolutionary Algorithms were proposed for the first time by Holland in 1975 [1]. The principal problems in the field of the Evolutionary Algorithm fund on the management of the compromise exploration/exploitation during the search for optimal solutions for optimization problem. It is by preserving diversity that the Evolutionary Algorithm can manage this compromise as well as possible. Diversity can be preserved by various ways, by controlling the pressure of selection, by seeking the good values for the parameters of the algorithm or by modifying the basic behavior of an Evolutionary Algorithm being inspired from natural phenomena.

Taking on consideration many references we already find that several models of Evolutionary Algorithm [2,3] are inspired from natural phenomena (model in islands, structuring of the populations, *etc*). Each of these models tries by applying its specific process to manage the compromise exploration/exploitation. They are particularly adapted for the optimization of some problems.

The algorithm that we will present is also inspired from a natural phenomenon, the relation "host-parasites" [4]. The main characteristics of this algorithm lay on the first hand on the conservation of some diversity and on the other hand on a storage capacity. These characteristics are necessary to optimize dynamic problems, where we have to follow a global optimum that varies depending on time parameter.

References

- Holland, J.H.: Adaptation in Natural and Artificial Systems. 2nd edition edn. MIT Press (1992)
- Cohoon, J.P., Hegde, S.U., Martin, W.N., Richards, D.: Punctuated equilibria: a parallel genetic algorithm. In Grefenstette, J., ed.: ICGA'87, Mahwah, NJ, USA, Lawrence Erlbaum Associates, Inc. (1987) 148–154
- Gorges-Schleuter, M.: Asparagos an asynchronous parallel genetic optimization strategy. In Schaffer, J., ed.: ICGA'89, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1989) 422–427
- Combes, C.: L'art d'être parasite. Les associations du vivant. Champs Flammarion (2003)

A Composition Mechanism Applied to Business Models

Térence Ferut

Laboratoire I3S, Université de Nice - Sophia Antipolis, France terence.ferut@gmail.com

The concepts proposed by object-oriented programming (encapsulation, polymorphism, modularity ...), represented a major progress for the construction of complex applications. They support the re-use and thus allow a major reduction in time of development and maintenance. It is however proven that the tools provided by these approaches are far from being sufficient to meet all the needs in re-usability.

The need to fill the gaps of the objects paradigm is the origin of many research works. It is particularly important to increase the adaptability of the software with an aim of making it evolve or developing it.

The re-use of classes, libraries, models of application or functional code falls under this step of evolution.

The approach developed by Laurent Quintian in his thesis [1] and improved in [2] is based on the programming by subjects and aspects, to regard an application as a set of concerns to be composed. One of the original aspects of this work was to use an approach driven by the models to implement the composition of these concerns.

The objective which was laid down to me is to transpose the ideas related to the composition of the concerns at the application level, from work of Laurent Quintian, to the composition of models. In other words, it is a question of adapting the process of models adaptations similar to the process of applications adaptation.

References

- 1. Quintian, L.: Un modèle pour améliorer la réutilisation des préoccupations dans le paradigme objet. PhD thesis, Université de Nice-Sophia Antipolis (2004)
- 2. Lahire, P., Quintian, L.: New perspective to improve reusability in object-oriented languages. Journal of Object Technology 5 (2006) 117–138

Contextual Spatial Ontology for Interoperability of Geographic Web Services

Florian Prud'homme

Laboratoire I3S, Université de Nice - Sophia Antipolis, France florian.prudhomme@gmail.com

Geographic information systems are an increasing key-factor in decision planning in various domains. However, efficient access to these systems is still an open problem. Relevant geographic information retrieval cannot be achieved through classical keywords matching algorithms. Actually, interesting geographic features are not stored, but computed by geographic servers. System integrators have to understand and learn thick API specifications for each geographic server, since functionalities are tied strongly with basic content and schema database. Manual integration is not very scalable.

We propose to exploit state-of-art semantic web system to achieve automatic geographic system integration. We will expose decidable logics and languages candidates and a roadmap we have designed to build formal descriptions enabling fully-automated integration.

Identifiability and Separability by Particle Filtering

Benoît Lagadec

Laboratoire I3S, Université de Nice - Sophia Antipolis, France blagadec@i3s.unice.fr

Some problems are quite common in communication systems. The overspecified case occurs if we try to provide an estimated signal given more observations than sources. On the opposite, the parsimonious case is an underspecified instance where all the sources are not present at the same time (consider for example a meeting where all the interlocutors do not speak at the same time).

The issue we are most interested in is to rebuild a source knowing both its distribution and its observation through a mixing matrix. This issue is known as the source separation problem. The number of sensors is often lower than the number of sources, and in this case we speak of prediction, rather than smoothing—where the number of sensors is higher than the number of sources— or filtering—where there is the same number of sensors and sources.

We first study under which conditions the prediction problem has a solution, that is when we have an identifiable system. Then we propose an algorithm that solves this problem according to various source distributions. The case of discrete sources differs slightly, for the estimated particles cloud does not always have the same distribution as the source.

Thereafter, we model the speech signal as Markov chains, *i.e.* correlated processes where the state at moment K depends on the previous states. We treat the case where the mixing matrix is unknown—the identification problem—and we are especially interested in convergent particle filters for an initial disturbance, *i.e.* for an *a priori* knowledge of the shape of the mixing matrix.

We then extend particle filtering to a convolutive case, which could represent a speaking signal. We can therefore give a solution to the infamous "cocktail party" problem. Note however that this case requires to multiply the states in the algorithm, and thus significantly increases the size of the mixing matrix. In the linear case of filtering with a state following a Gaussian noise, there is an exact solution: in 1961, Kalman and Bucy introduced a filter with an essential interest: it is recursive and can easily be implemented on a computer.



Les séminaires doctorants

Les séminaires des doctorants STIC permettent aux futurs docteurs d'échanger leurs expériences dans leur travail de thèse, tant sur le plan scientifique que sur le plan professionnel et éducatif. Ces rencontres ont lieu mensuellement dans l'un des laboratoires STIC de Sophia Antipolis.

Un séminaire est l'occasion de trois à quatre interventions, dont une effectuée par un jeune permanent. Chaque intervention comporte un exposé technique d'une vingtaine de minutes et une période d'échanges et de retours d'expérience d'une dizaine de minutes.

Ces actes compilent les résumé en anglais des exposés techniques du séminaire doctorant du 14 juin 2006.

L'ADSTIC

L'ADSTIC est l'association des doctorants du campus sciences et techniques de l'information et de la communication de l'université de Nice Sophia Antipolis. Créée en 2004, l'ADSTIC est une association loi 1901.

Notre but essentiel est de faciliter les contacts entre les doctorants des différentes disciplines présentes sur le campus STIC, de les informer et de valoriser leur formation doctorale. L'ADSTIC se veut aussi un lien entre les doctorants passés, actuels et futurs...

Pour plus de renseignements, visitez notre site Internet : http://adstic.free.fr.