

Fiches d'activité des doctorants

Université de Lyon Lumière Lyon 2
5, avenue Pierre Mendès-France
69676 Bron Cedex

Tel. +33 4 78 77 23 76

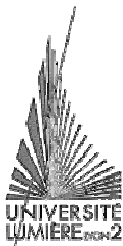
Fax. +33 4 78 77 23 75

Web. <http://eric.univ-lyon2.fr>

Université de Lyon, C. Bernard Lyon 1
43, bd du 11 novembre 1918
69622 Villeurbanne cedex

Tél. +33 4 72 43 16 54

Fax +33 4 72 43 10 44



Ecole Doctorale
Informatique et Mathématiques de Lyon
(InfoMaths)

Emna BAHRI

Current Position : PhD student
E-mail : Emna.bahri@univ-lyon2.fr
Web site : Eric.univ-lyon2.fr/ebahri
Birth Date : 15/04/1981
Arrival Date : 17/10/2006



Research supervisor : Stéphane Lallich

Research topics

In spite of a great number of theoretical and practical results on various fields, machine learning must nevertheless deal with certain problems of applicability on real data which are generally complex. Most known is the treatment difficulty of large volumes of data, and the intolerance with the presence of noise. These problems always constitute a research topic to improve the training process.

In fact, my research aim is to improve the training methods in order to have a powerful prediction systems i.e. to obtain a classifier having a good (but seldom perfect) success rate, but does not matter with data nature (noised or not). We can divide our prospects for the thesis in two parts according to problems of the noise and the imbalanced data. The first prospect was realized throughout the first year. In fact, we proposed a new approach called AdaboostHybride, an improvement of boosting against the noise. This approach aims exploiting assumptions generated with the former iterations of AdaBoost to act both on the modification of the weight of examples and the modification of the classifiers' weight. The second prospect is based on the imbalanced data problem. We choose for this problem the recent approach of classification based on rules called the associative classification. Since the first work of (Liu, Hsu, and Ma 1998), various works show the good performance of associative classification in terms of error rate reduction. However, these methods have various weaknesses which our work proposes to improve. In fact, we propose W-CARP an associative classification method. The purpose of this new method is to give results that are at least equivalent to those of the existing approaches but with an execution time much more reduced. To achieve the first phase, that of the class rules extraction, W-CARP use FCP-Growth-P, an adaptation of FP-Growth. FCP-Growth-P builds only the class itemsets. In fact, it uses a support threshold adapted to each class and a procedure of pruning. To carry out the second phase, we propose to build a rule's base which groups the most significant rules and to weigh the predictions of the various rules covering an example by the value of their Loevinger measurement. Actually, we try to improve the quality of prediction using the Boosting. Our new approach is called CARBOOST.

Publications

- E. Bahri, S. Lallich.(2009): Improving prediction by weighting class association rules, The fifth International Conference on Machine Learning and Application, Miami, Florida, USA.
- E. Bahri, S. Lallich.(2009): Pruning for extracting association rules without candidate generation, The fifth International Conference on Data Mining(DMIN 09), Las Vegas, Nevada, USA
- E. Bahri., S.Lallich (2009): Introduction de l'élagage pour l'extraction de règles d'association de classe sans génération de candidats, Atelier Qualité de données et de connaissances (EGC09), Strasbourg, France, (2009).
- E. Bahri, S. Lallich, N. Nicoloyannis, M. Maddouri (2008), «A hybrid Approach of Boosting against noisy data», in Mining Complex Data, Ras Z. W., Tsumoto S. and Zighed, D. A. (Eds.), Lecture Notes in Artificial Intelligence,vol. 4944, Springer.
- E. Bahri, M. Maddouri (2008), « Nouvelle approche du Boosting pour les données bruitées », (EGC 08), Nice, Janvier 2008.

Soumaya BEN HASSINE GUETARI

Current Position : PhD student
E-mail : sbenhassine@aid.fr
Web site :
Birth Date : 08/11/1982
Arrival Date : 01/10/2008



Research supervisor : Jérôme Darmont & Jean-Hugues Chauchat

Research topics

Enhancement of data quality in Databases and Data Warehouses: achievement of the concept “introspective system” and application to the targeting operations in the marketing Databases.

Our aim is to introduce an information system that:

- handles the quality of a multi-source information system: that is able to act and react in order to maintain the quality of its data; and which is able, therefore, to select the best source (containing the most accurate and complete data)
- handles the BtoB process and provides the data needed to analyse and establish strategic decisions regarding to the targeting and marketing operations (basing on the ROI criteria, the customer profile, objective data quality dimensions...).

Publications

S. Ben Hassine, D. Clément, B. Laboisse. « Using association rules to detect data quality issues». In proceedings of the 13th ICIQ (MIT). Novembre 2008.

S. Ben Hassine, D. Clément. « Règles d’association pour la qualité des données ». QDC 09. Janvier 2009.

S. Ben Hassine “Data quality evaluation in an e-business environment : a survey”, In proceedings of the 14th ICIQ conference, Berlin, 2009.

Sonia BOUATTOUR

Current Position : PhD student
E-mail : sonia.bouattour@eric.univ-lyon2.fr
Web site : <http://recherche.univ-lyon2.fr/eric/85-Sonia-BOUATTOUR.html>
Birth Date : 03/07/1983
Arrival Date : 01/10/2007
Research supervisor : Omar Boussaid, Jamel Feki (MIRACL – Sfax)



Research topics

Architecture definition of an active data warehouse for the automation of analysis scenario.

Data warehouses were introduced in the 90s. Since then, they have known a widespread acceptance and an important progress toward an efficient support for decision making. In fact, data are usually designed according to multidimensional structures, commonly called data cubes. Each data cube expresses a specific analysis context that may be visualized, explored and summarized thanks to OLAP operators. Analytical reports can therefore be generated from multidimensional data according to the needs defined by the business decision makers. Nevertheless, traditional decision systems are not equipped with automatic analysis mechanisms and can not maintain dynamic OLAP reporting.

Therefore, we propose an approach of an active data warehouse (ADW). We particularly exploit the notion of active rules already known in active databases. An active rule consists traditionally of events, conditions and actions. This triplet can be used to fully define analysis rules that express scenarios used for OLAP reporting. Thus, our proposition consists in integrating these analysis scenarios into the decision system regarding events, conditions and actions that can express usual analysis needs of enterprises.

We propose a simplified and general formalization of an active data warehouse framework. By exploiting the ECA formalism known in active databases, we introduced a new formalism of analysis rules and defined therefore a general framework of an active data warehouse. In this framework, XML is used to model both the logical and the physical level of analysis rules. We also provide an software prototype in order to validate our approach.

Publications

S. Bouattour, O. Boussaid, *Les entrepôts de données actifs : modélisation des règles d'analyse*, In 5emes Journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA' 09), Montpellier, Juin 2009, Cepadues Éditions, Toulouse.

S.Bouattour, R.Ben Messaoud, O.Boussaid, H.Ben Abdallah, J.Feki. A Framework for Active Data Warehouses. *The 2008 International Arab Conference on Information Technology (ACIT'2008)*, Hammamet, Tunisie, 16-18 Décembre 2008.

S.Bouattour, J.Feki, H.Ben Abdallah, R.Ben Messaoud, O.Boussaid. Proposition d'une taxonomie d'événement dans les entrepôts de données actifs. *Les huitièmes journées scientifiques des jeunes chercheurs en génie électrique et informatique (GEI'2008)*, Sousse, Tunisie. 17–19 Mars 2008.

S. Bouattour, R.Ben messaoud, O. Boussaid, "Modélisation de règles d'analyse dédiées aux entrepôts de données actifs", 2^{ème} édition de l'atelier des systèmes décisionnels (ASD 07), Sousse, Octobre 2007.

Sonia Bouattour, « Les entrepôts de données actifs : automatisation des scénarii d'analyse », Mémoire de Master, Université Lumière Lyon 2, Lyon, France. Juin, 2007.

Mathilde FORESTIER

Current Position : PhD student

E-mail : Mathilde.forestier@univ-lyon2.fr

Web site : <http://recherche.univ-lyon2.fr/eric/25-Mathilde-Forestier.html>

Birth Date : 09/09/1985

Arrival Date : 01/10/2008

Research Director : Djamel Zighed

supervisor : Co-supervisor : Julien Velcin



Research topics

The Web 2.0 contains a lot of data related to significant information. The aim of my Phd research is to extract social networks from the Web and to build a model of interactions including the opinion. We plan to use machine learning and data mining techniques to automatically extract social networks and people's opinion. Several scientific obstacles must be overcome, such as the detection of named entities, their relations but also the detection of the very nature of these relations.

For example, in a forum, we would like to extract the people who speak about the same subject. Who disagrees with who? Who is the leader of the discussion? Etc.

This kind of research begins to appear in the international world research. It exists two significant previous works dating from 2004¹ and 2007² that extract automatically social networks. The particularity of my research is to include in my model the opinion and the topics that people are talking about.

Even if the forums are our primary material, we would like to extend the model to other data sources and other forms of opinion content from around the Web (blogs, chats...).

Publication

M. Forestier, J. Velcin, J. Ganascia, "Un cadre formel pour la veille numérique sur la presse en ligne", Atelier Veille Numérique (EGC-VN 09), Strasbourg, Janvier 2009.

¹ Culotta, A. and Bekkerman R. and McCallum A., Extracting social networks and contact information from email and the Web, First Conference on Email and Anti-Spam (CEAS), 2004

² Jing, H. and Kambhatla, N. and Roukos, S., Extracting social networks and biographical facts from conversational speech transcripts, ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, vol.45,2007

Quentin GICQUEL

Current Position : Phd student
E-mail : Quentin.gicquel@gmail.com
Web site :
Birth Date : 12/11/1985
Arrival Date : 01.2009



Research supervisor : Stéphane Lallich ; Marie-Hélène Metzger

A nosocomial infection is an infection that is said in the hospital 48 hours after admission.

Base work: I work on 2 bases. The first relates to patients in intensive care unit (ICU) in hospices civils de Lyon. It includes approximately 1500 persons and 200 variables. These variables were derived from several departments (ICU, biology, diagnosis, drugs ...). We created 5 bases for each site of nosocomial infection.

The second basis is that of a project ANR. This project is to work directly on the medical records of ICU and surgery (3 different services). This database is not yet finalized. There will be more variables (history) and more individual (2000).

Method: We only worked on first base. After having built it, we used several methods Data mining: Decision tree, random forest, predictive association rules, rules boosted, SVM, logistic regression...

Results: Except for the predictive association rules, we obtain the confusion matrices. From there, we can study several indicators (error rate, specificity, accuracy ...). They give us very poor prediction rate of infection, but very well for predicting healthy. This trend is accentuated when there are few infections. This reflects the fact that initially there are few of infection. Even if the risk is increased by 3 or 4, sometimes (depending on the site) that the infection is still a minority. Regarding the association rules, we find the same trend: few rules with low support and low confidence. A few are interesting in terms of lift.

Outlook: Our main perspective is to use known methods when the imbalance of classes (including the class to predict) or otherwise to create one from the known. Therefore, we believe work on the training set, on the functions of bursts of nodes, on the prediction rule in groups, on the validation process...

Publications

- **Q Gicquel**, MH Metzger, A Lepape, E Bahri, S Lallich : " Risk factor analysis for nosocomial infections in intensive care unit : a data mining approach." 9th International Conference on System Science in Health Care, Septembre 2008.
- Metzger MH, **Gicquel Q**, Proux D, Pereira S, Kergourlay I, Serrot E, Segond F, Darmoni S. Development of an automated detection tool for healthcare-associated infections based on screening natural language medical reports. AMIA Annu Symp Proc. 2009
- Proux D, Marchal P, Segond F, Kergourlay I, Pereira S, **Gicquel Q**, Darmoni S, Metzger MH. Improving Hospital Document workflow with a Risk Patterns Detection tool to detect potential Hospital Acquired Infections. International Conference on Recent Advances in Natural Language Processing. Borovets, Bulgaria 14-16 september, 2009.

Oksana GRABOVA

Current Position : PhD student
E-mail : oksana.grabova@eric.univ-lyon2.fr
Web site : <http://recherche.univ-lyon2.fr/eric/111-Oksana-Grabova.html>
Birth Date : 27/12/1982
Arrival Date : 01/02/2009
Research supervisor : Pr. Jérôme Darmont, Pr. Jean-Hugues Chauchat



Research topics

Web data warehouses (for small and middle-sized organizations).

During the last years data warehouse systems have become an essential component of modern decision support systems in most companies of the world. In order to be competitive even small and middle-sized organizations now collect large amounts of information and are interested in business intelligence systems. In spite of multiples advantages of existing decisions decision support systems (DSS) remains frequently inaccessible or insufficient for SME (small and middle-sized enterprises) because of following factors: high price, high requirements for a hardware and infrastructure, complexity for most users, nonoptimal correlation of cost-effectiveness, redundant functionality, no flexibility enough to deal with a fast changing dynamic business environment, don't pay attention to a big difference at necessity of data access in SME and large-scaled enterprise.

Moreover, many decision-support applications necessitate external data sources. Besides, there is a clear trend toward on-line data warehousing, which gives way to new approaches such as virtual warehousing or XML warehousing.

Thus, the PhD applicant's objective is to formalize and design simple, lightweight, on-line data warehouses with user-friendly interfaces.

Publications

Marouane HACHICHA

Current Position : PhD student
E-mail : marouane.hachicha@eric.univ-lyon2.fr
Web site : <http://eric.univ-lyon2.fr/~mhachicha/>
Birth Date : 07/June/1983
Arrival Date : 01/October/2007
Research supervisor : Prof. Jérôme Darmont



Research topics

The objective of this thesis is to allow OLAP analysis of complex data structured in XML data warehouses. This work consists on the design of a XML-OLAP (or XOLAP) algebra in the order to carry out traditional OLAP queries on native XML data.

This new formal framework represents the first step of our work. Then, it will be necessary to enrich this XOLAP algebra with new specific operators to the XML context. Then, it appears necessary to be able to carry out some operations like roll up and drill down on complex hierarchies of dimensions such as the ragged hierarchies. Operators coupling the principles of OLAP and Data Mining could also allow the treatment (aggregation) of multimedia data resulting from the Web. This work also aims at supporting the efforts of the extension of the XQuery language for the decisional applications.

In addition, to have a XOLAP algebra for the decisional complex data processing must allow the optimization of OLAP queries expressed in XQuery. Native XML Data Bases Management Systems (DBMS), although in a constant progress, present some limitations in term of performance and would profit largely from an automatic queries' optimization, particularly the decisional queries which are, generally, very expensive.

Finally, an implementation of this work on XOLAP is envisaged within the framework of a platform of complex XML data storage, under development at the ERIC laboratory. The objective is to allow, through a simple and accessible interface since the Web, the construction and the handling of XML cubes of complex data.

Publications

1. Hadj Mahboubi, Marouane Hachicha, Jérôme Darmont, "XML Warehousing and OLAP", Encyclopedia of Data Warehousing and Mining, Second Edition, IGI Publishing, Hershey, PA, USA, 2008.
2. Marouane Hachicha, Hadj Mahboubi, Jérôme Darmont, "Expressing OLAP operators with the TAX XML algebra", 3rd International Workshop on Database Technologies for Handling XML Information on the Web (DataX-EDBT 08), Nantes, France, March 2008.
3. Marouane Hachicha, Hadj Mahboubi, Jérôme Darmont, "Vers une algèbre XML-OLAP : état de l'art", 2^{ème} Atelier Systèmes Décisionnels (ASD 07), Sousse, Tunisie, Octobre 2007.

Rym KHEMIRI

Current Position : PhD student
E-mail : Rym.Khemiri@univ-lyon2.fr
Web site :
Birth Date : 06/07/1984
Arrival Date : 18/09/2009



Research supervisor : Fadila Bentayeb

Research topics

Le but de cette thèse est la spécification d'un ensemble d'outils d'aide à la personnalisation des entrepôts de données.

Les travaux sur la personnalisation dans les entrepôts de données ont consisté principalement à étendre les approches habituelles utilisées principalement dans les domaines des Bases de Données (BD) et de la Recherche d'Information (RI) en permettant de modifier les structures multidimensionnelles et/ou les mécanismes de présentation des données par une meilleure connaissance des utilisateurs. Pour ce faire, ces travaux exigent des efforts cognitifs de la part de l'utilisateur qui doit exprimer de manière explicite les préférences qui le caractérisent. Plus précisément, nous traçons trois axes majeurs de recherche.

Un premier axe concerne la construction de profils utilisateurs dans les entrepôts de données. Cet aspect a été largement étudié notamment en RI, mais, à l'heure actuelle, aucune solution unanimement reconnue ne semble émerger.

Pour remplir cet objectif, il est nécessaire de définir des solutions permettant l'évolution des profils, leur extensibilité. Les caractéristiques de ces derniers sont constituées principalement des préférences utilisateur sur les données et les structures de l'entrepôt.

Un deuxième axe porte sur l'influence de la personnalisation sur la modélisation multidimensionnelle. Dans ce contexte, on peut effectuer le couplage de techniques de fouille de données et de la modélisation multidimensionnelle pour inférer des hiérarchies à recommander.

Un troisième axe d'étude repose sur la définition de structures de visualisation alternatives à la visualisation traditionnelle sous forme de table à deux dimensions. La problématique est alors d'inclure le processus de personnalisation par rapport à la visualisation. L'objectif sera alors de prendre en compte, dans la définition et l'utilisation de structures de visualisation, des caractéristiques propres à l'utilisateur, par exemple, son niveau d'expertise.

Publications

- Khemiri, R et Akaichi, J. Memorizing and Analyzing Floods Trajectory. 3rd. International Conference on Information Systems and Economic Intelligence (SIEE'2010). 18-20 Février 2010 Sousse, Tunisie.

Nora MAIZ

Current Position : PhD student
E-mail : nmaiz@eric.univ-lyon2.fr
Web site : <http://eric.univ-lyon2.fr/~nmaiz>
Birth Date : 08/04/1979
Arrival Date : 01/11/2005



Research supervisor : Pr. Omar BOUSSAÏD and Dr. Fadila BENTAYEB

Research topics

Actually, OLAP (On Line Analysis Process) is performed using voluminous multidimensional databases named data warehouses. Data is integrated from heterogeneous and distant data sources to be loaded according to the data warehouse model, in a uniform format. Mediator consists in integrating data using a virtual layer, which gives a global view on different data sources. However, it is not oriented on line analysis. We aim in our work to define a global on line analysis strategy using mediator approach. The use of ontologies in the mediation process allows semantic and structural integration. We propose so a new mediation system based on an hybrid architecture of ontologies modelled according to GLAV (Generalized Local As View) model. The hybrid architecture builds a local ontology for each data source and a global ontology viewed as a mediator. The integration model defines how sources, local and global ontologies are mapped. We propose so an ascending method for building ontologies, which facilitates the semantic reconciliation between data sources. Moreover, we use OWL (Ontology Web Language) for defining ontologies and mappings between data sources and ontologies. In first time, building global ontology is manual. After that, we propose an ontology merging strategy based on hierarchical clustering. User queries are expressed in our specific language which handles global ontology concepts and local ontologies properties because we assume that the user is expert in its domain. Queries are decomposed by the rewriting algorithm to obtain a set of equivalent sub-queries that are sent to the corresponding sources for execution, and after that recomposed to obtain the final result.

Publications (5 maxi)

[MBBa] N. Maiz, O. Boussaïd, F. Bentayeb, "Ontology-Based Mediation System", 13th ISPE International Conference on Concurrent Engineering: Research and Applications (CE 06), Antibes, France, September 2006 ; Frontiers in Artificial Intelligence and Applications, Vol. 143, 181-189.

[MBBb] N. Maiz, O. Boussaïd, F. Bentayeb, "A Hybrid Architecture of OWL-Ontologies to Relational Data Sources Integration and Query Treatment", 18th Information Resources Management Association (IRMA07). International Conference on Managing Worldwide Operations and Communications with Information Technology. May 19-23, 2007, Vancouver, Canada

[MBBc] N. Maiz, O. Boussaïd, F. Bentayeb, "Ontology Merging by Clustering For Data Warehouse Building on-the-fly". SFC-CLADAG, Juin 2008. Italie.

[MAD] N. Maiz, K. Aouiche, J. Darmont, "Sélection automatique d'index et de vues matérialisées dans les entrepôts de données", 2ème journée francophone sur les Entrepôts de Données et l'Analyse en ligne (EDA 06), Versailles, 2006 ; Revue des Nouvelles Technologies de l'Information, Vol. B-2, 89-104.

[MBBd] N. Maiz, O. Boussaïd, F. Bentayeb, "Fusion des ontologies par classification hiérarchique", 2ème journée francophone sur les ontologies (JFO08), Lyon, 2008.

Efthimios MAVRIKAS

Current Position : PhD student
E-mail : efthimios.mavrikas@eric.univ-lyon2.fr
Web site : -
Birth Date : 04/03/1979
Arrival Date : 06/01/2003
Research supervisor : Djamel. ZIGHED



Research topics

Cultural Heritage documents deal with objects/artifacts and the people that created, owned, used, or (re)discovered them. Their fates are intertwined in unique and complex stories forming a cumulative body of knowledge, often fragmented across large online document collections. While our collective memory has explicitly documented these stories, the heterogeneity of the available sources creates islands of information that can only be implicitly connected by a limited, expert audience.

My current research work aims to define a semantically consistent framework for the online presence of Cultural Heritage document collections, set upon a participatory centre stage and supported by a shared knowledge model. In this framework, Cultural Heritage document contributors benefit from knowledge-rich document processing modules which analyse and classify each contribution, capture the notion of time and the unfolding of events spanning single or multiple documents, and establish meaning connectivity over the entire collection. Overall, this framework assists a scholarly audience with the exploration of online Cultural Heritage document collections, and offers an informed tap into the collective memory scattered therein.

Keywords: Discourse Analysis, Ideology, CIDOC CRM, WorldNet, PLSA, Scripts.

Publications

Efthimios C. Mavrikas, Evangelia Kavakli and Nicolas Nicoloyannis (2005) The Story between the Lines: Exploring Online Cultural Heritage Document Collections using Ontology-based Methods, *Annual Conference of the International Committee for Documentation of the International Council of Museums (CIDOC 2005)*, Zagreb, Croatia, May 2005.

Efthimios C. Mavrikas, Vagelis Stournaras and Christis Konnaris (2005) Historical Memory Preservation on the Semantic Web: the Case of the Historical Archive of the Aegean - Ergani, *33rd International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA 2005)*, Tomar, Portugal, March 2005.

Efthimios C. Mavrikas, Evangelia Kavakli and Nicolas Nicoloyannis (2004) Ontology-based Narrations from Cultural Heritage Texts, *5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2004)*, Ename, Belgium, December 2004.

Efthimios C. Mavrikas, Nicolas Nicoloyannis and Evangelia Kavakli (2004) Cultural Heritage Information on the Semantic Web, *14th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2004)*, Northamptonshire, UK, October 2004, Springer LNAI, vol. 3257, pp. 477-478.

Dimitris C. Papadopoulos and Efthimios C. Mavrikas (2003) Peer-to-Peer Ways to Cultural Heritage, *31st International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA 2003)*, Vienna, Austria, April 2004.

Huu Hoa NGUYEN

Current Position : PhD student
E-mail : nhhoa@eric.univ-lyon2.fr
Web site : <http://eric.univ-lyon2.fr/~nhhoa/>
Birth Date : May 24, 1973
Arrival Date : October 1, 2009



Research supervisor : Nouria HARBI & Jérôme DARMONT

Research topics

During the past decade, intrusion detection has been a widely studied topic. Many researchers have endeavored to propose various intrusion detection approaches (i.e. artificial neural network, fuzzy set, evolutionary computation, artificial immune, and data mining based approaches). These approaches have contributed to intrusion detection problems in different ways together with their own pros and cons. Although several promising results have been achieved, existing researches have not been mature enough to fit the demand of implementing a good intrusion detection system. This is due to the fact that there remain many challenging aspects involved in intrusion detection. Some of them can be recognized as follows. Firstly, data for analysis is extremely massive, noisy, and complex. Moreover, availability of benchmark datasets for training and evaluation is limited, though there exists an outdated and arguable benchmark KDD' 99 dataset. Secondly, for anomaly detection, normal behaviors keep evolving constantly and hence it is challenging to identify boundary between normal and anomalous behaviors. Lastly, detecting intrusions in practice requires not only high accuracy and fast performance but also less human intervention. These characteristics require efficient algorithms and strong methods. In our view, data mining is a promising research direction for intrusion detection problems. More particularly, assembling different methods in cooperative ways definitely intensify the effectiveness of intrusion detection system.

The main objective of my thesis is to focus on the issues of network intrusion detection based on data mining approaches to handle shortcomings of current intrusion detection solutions. Our efforts are to devise efficient algorithms and robust methods for improving detection rate and performance while controlling false alarm rate at an acceptable level. We plan to propose hybrid mechanisms that can efficiently detect both foreseen and unforeseen attacks by combining the preeminence of misuse-based and anomaly-base analysis techniques. Another concentration is to investigate appropriate techniques to deal with the complexity of network traffic data sources and the constant change of network system environments with respect to both intrusive and normal behaviors.

Publications

D.M. Farid, J. Darmont, N. Harbi, H.H. Nguyen, M. Rahman, "Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification", International Conference on Computer Systems Engineering (ICCSE 09), Bangkok, Thailand, December 2009.

Vincent PISETTA

Current Position : PhD student
E-mail : vpiSETTA@eric.univ-lyon2.fr
Web site :
Birth Date : 20/04/1982
Arrival Date : 19/08/2008



Research supervisor : Djamel A. Zighed

Research topics

A lot of techniques have ever been developed aiming at reducing the number of features in a dataset, i.e., reducing the dimensionality of data (for example principal component analysis). In this thesis, we will focus on approaches assuming that this high dimensionality is only artificial and that features are actually measurements of the same underlying cause. For example, gray scale $n \times n$ images of a fixed object taken with a movie camera yield data points in a space of dimension n^2 . However, the intrinsic dimensionality of the space of all images of the same object is the number of degrees of freedom of the camera. In fact, the space has a natural structure of manifold embedded in a space of dimension n^2 . Several algorithms such as Laplacian Eigenmaps (Belkin, 2003), ISOMAP (Tenenbaum, 2000) or LLE (Roweis and Saul, 2000) and more generally topologically-based approaches have been developed in order to find such manifolds.

All those algorithms require an input in the form of a similarity matrix between instances. While there is a lot of work concerning the way of properly identifying manifolds given a similarity matrix, a few attentions have been dedicated to the construction of the similarity matrix itself. In this thesis, we study the construction of similarity matrix properly designed for supervised and unsupervised learning tasks. Our framework is based on the construction of an ensemble of pseudo-randomized trees. In the purely supervised case, random forest (Breiman, 2001) is particularly studied as a kernel (similarity) designer and we study its infinity behavior.

Publications

Pisetta, V., Jouve, P.E., Zighed, A.D (2010). Construction de noyaux pour l'apprentissage supervisé à partir d'arbres aléatoires, 10eme Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances, EGC 2010, Hammamet, Tunisie.

Pisetta, V., Zighed, A.D (2009). Similarity and Kernel Matrix Evaluation based on Spatial Autocorrelation analysis, 18th International Symposium on Methodologies for Intelligent Systems, ISMIS, Prague.

Pisetta, V., Zighed, A. D (2009). Kernel Construction via Graph Adaptation, 1st international workshop on Topological Learning, Prague.

Pisetta, V., Zighed, A. D (2009). L'effet du voisinage sur la séparabilité des classes : application à la selection de noyaux pour les SVM, 1st workshop of Topological Learning, Starsbourg, p.25-33.

Taimur Qureshi

Current Position : PhD student
E-mail : taimur.q80@gmail.com
Web site : NA
Birth Date : 10/09/1980
Arrival Date : 01/10/2006
Research supervisor : D.A.Zighed

Research topics

The goal of this PhD can be divided into two parts:

1) Practical Part:

- Design and implementation of a generalized algorithm for decision trees and induction graphs.
- Development of software that incorporates the above algorithm and thus, creation of a test bed for experimentation using vast data and understanding of various algorithms and techniques.
- Development of an internet based collaborative tool that implements a huge data store for decision trees and induction graph based resources e.g. articles with their summaries, tools, books etc.

2) Theoretical Part: This part concerns with the development of new techniques and methods in the area of decision trees and experimentation with huge data on the created test bed and obtaining applicable results.

1) Practical Part:

In this portion a generalized decision tree and induction graph algorithm has been conceived and designed using flowcharts and object oriented designing techniques. The concept of the generalized algorithm is to create such an algorithm which is generic and can be used to implement anyone of the existing decision tree or induction graph techniques. We have implemented various discretization algorithms such as Chi merge, FUSBIN, FUSINTER, MDLPC, CONTRAST and also various decision trees as ID3, C4.5, CART and Arbogodai. We have implemented our algorithm in R and once the object oriented implementation is completed, we will transfer it into our software which is implemented in Java. The software has a user friendly interface that converts many types of data e.g. text, xml, db etc into a table structure. After that the user can select the type of technique to use on that data and the results shall be given as a graphical output. The third phase of the implementation is development of an internet based collaborative platform for decision tree and induction graph based resource sharing. We have developed a "Wikipedia" like tool for information sharing and editing. It shall contain resumes and sources of many articles, tools and platforms and a test bed; thus forming a complete resource for decision tree and induction graphs.

2) Theoretical Part:

From various studies done earlier, we know that the learning sample is an approximation of the whole population, so the optimal discretization built on a single sample set is not necessarily the global optimal one. Whereas, we proved that resampling gives a better estimate of the discretization point distribution in terms of achieving a well-defined distribution. We have created a discretization point selection protocol which selects cut points from a certain frequency distribution achieved by resampling. This protocol selects the discretization points from a given frequency point distribution having higher probability of occurrences and splits on those points if a certain criterion (e.g. entropy) is met. When we apply that protocol, it significantly improves the quality of discretization and prediction rate and thus, nearing to a global optimal solution. Moreover, the same protocol when applied to the frequency point distribution of random samples, achieved much lesser improvements in the prediction rate as compared to bootstrap. We applied the discretization point selection protocol (after resampling) to various methods on the breiman waveform dataset. Except for Chi-Merge, all the other methods provide small variations in terms of prediction rates. MDLPC performs the best and FUSBIN achieves the best time complexity, which is a key point when dealing with a lot of examples.

We applied the above mentioned resampling methodology in the context of fuzzy or soft discretization in decision trees. Our soft discretization technique gives better prediction rates than the hard discretization based methods. As ongoing work, we are applying this soft discretization in building soft decision trees and thus, will try to show that this method will also improve the classification accuracy of decision trees.

Publications

IEEE ICSEA-2004: Integration of Mobile IP and Adhoc Networks with Multi-homing and Smooth Handoff capabilities.

IEEE 16th IST Mobile Summit, Budapest, Hungary: A Network Layer based Hard Real Time Protocol for Wireless Sensor Networks.

Jean-Christian RALAIVAO

Current Position : PhD student
E-mail : jean-christian.ralaivao@eric.univ-lyon2.fr
Web site : <http://eric.univ-lyon2.fr/~jcralaivao/>
Birth Date : 03/07/1966
Arrival Date : 01/11/2003
Research supervisor : Jérôme Darmont & Stéphane Lallich

Research topics

Within decision processes, data warehousing technologies are now mature to handle simple, numerical or symbolic data. However, various sources including the Web store very heterogeneous data: texts, images, sounds, videos, databases, temporal or geographical data; which may be expressed in several languages, stored in various formats, located in different places and frameworks, etc. These so-called complex data carry a lot of information and are thus interesting to include within a decision process. However, numerous issues relate to structuring, storing and querying complex data.

The aim of my PhD thesis is to address the issue of complex data warehouse performance. Several techniques do exist to optimize simple data access and storage in a warehouse. However, they cannot apply very efficiently onto complex data. Thus, we have to define complex data warehouse models that are adapted to the nature of stored data, and to design custom performance optimization tools for these warehouses: indexing, view materialization, partitioning, clustering, buffering, etc.

Aside, using the XML language for managing data warehouses has several advantages, especially when integrating heterogeneous data. XML indeed helps represent both structure and contents. Hence, we have proposed an XML-based complex data warehouse architecture [1, 3] that helps benefit from optimization techniques developed in the database and XML communities.

Performance optimization always needs well-defined measures and metrics. In order to identify them, we listed performance indicators for complex data warehouses. This list helped identify performance factors that are used to determine metrics.

Finally, integrating metadata and domain-related knowledge in the complex data warehouse has a positive impact on managing data complexity, especially in the process of performance optimization [2, 4].

Publications

1. H. Mahboubi, J.C. Ralaivao, S. Loudcher, O. Boussaid, F. Bentayeb, J. Darmont, "X-WACoDa: An XML-based approach for Warehousing and Analyzing Complex Data", In L. Bellatreche, Ed., "Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction", *Advances in Data Warehousing and Mining*, IGI Publishing, Hershey, PA, USA, 2009.
2. J.C. Ralaivao, J. Darmont, "Knowledge and Metadata Integration for Warehousing Complex Data", *6th International Conference on Information Systems Technology and its Applications (ISTA 07)*, Kharkiv, Ukraine, May 2007; *Lecture Notes in Informatics (LNI)*, Vol. P-107, 164-175.
3. J. Darmont, O. Boussaïd, J.C. Ralaivao, K. Aouiche, "An Architecture Framework for Complex Data Warehouses", *7th International Conference on Enterprise Information Systems (ICEIS 05)*, Miami, USA, May 2005, 370-373.
4. J.C. Ralaivao, "Améliorer la performance d'un entrepôt de données complexes par l'utilisation de métadonnées et de connaissances du domaine", *2ème atelier Fouille de Données Complexes dans un processus d'extraction des connaissances, EGC 05, Paris*, Janvier 2005, 81-84.

Marian-Andrei RIZOIU

Current Position : PhD student
E-mail : Marian-Andrei.RizoIU@univ-lyon2.fr
Web site : <http://recherche.univ-lyon2.fr/eric/114-Marian-Andrei-RizoIU.html>
Birth Date : 31/03/1985
Arrival Date : 01/10/2009

Research supervisor : Director : Stéphane Lallich
Co-supervisor : Julien Velcin



Research topics

Numerous institutes have developed information systems dedicated to specialists in human and social sciences. These systems rely on heterogeneous (different means of storage), complex (containing text, fields, photos etc) and voluminous collections of data.

The purpose of this PhD thesis is to research for means of automatic organization of this data in order to render it usable for the researchers in sociology, history, political sciences etc. Our current approach is to structure the available information under the form of hierarchies, using :

- methods of **conceptual clustering** [Godoy et al, 2006] ;³
- the information existing before-hand under the form of tags or annotations applied on the data by the experts – **semi-supervised approach** [Gao et al, 2006] ;⁴
- methods of concept extraction which allow **concept overlapping** [Cleuziou, 2007].⁵

As applications of the PhD thesis research, we have considered :

- structuring data into historical events, generating an event hierarchy based on the tags provided by the historians ;
- structuring the topics that emerge from online discussions

Publications

M-A. RizoIU, J. Velcin, J-H. Chauchat, "Regrouper les données textuelles et nommer les groupes à l'aide de classes recouvrantes", Conférence EGC 2010.

³ D. Godoy, A. Amandi : "Modeling user interests by conceptual clustering. Information System", 31(4-5) : pp. 247-265 (2006)

⁴ J. Gao, P. Tan et H. Cheng : "Semi-supervised clustering with partial background information ", In Proceedings of the Sixth SIAM International Conference on Data Mining (2006)

⁵ G. Cleuziou : "Okm : une extension des k-moyennes pour la recherche de classes recouvrantes". In M. Noirhomme-Fraiture et G. Venturini (Eds.), EGC, Volume RNTI-E-9 of Revue des Nouvelles Technologies de l'Information, pp. 691–702. Cépaduès-Éditions, (2007)

Rashed SALEM

Current Position : PhD student
E-mail : rashed.salem@univ-lyon2.fr
Web site : <http://eric.univ-lyon2.fr/~rsaalem/>
Birth Date : 20/10/1981
Arrival Date : 01/11/2007
Research supervisor : Jérôme Darmont & Omar Boussaïd



Research topics

Warehousing data for decisional support purposes is becoming non trivial task, due to their complexity with the presence of a huge amount of distributed and heterogeneous data. Therefore, we exploit XML as pivot language in order to unify, model, and store complex data. In this context, developing an XML-based data warehouse for complex data involves two major phases: one of integrating it, by extracting, transforming, and loading (ETL) complex data into a specified XML repository, and the other of analyzing it, by applying eXtensible OLAP (XOLAP) techniques to better answer the user's queries. However, traditional XML data warehousing lacks of continuously up-to-date information, which results in poor decisions. Enriching the proposed warehouse with active rules and Web services yields it to be always up-to-date. In our thesis project, we focus on integration phase.

Main objective:

- Design, implement and validate an architecture for integrating complex data into an unified repository autonomously.

Challenges:

- Including semantics of complex data.
- Detecting events of changes in some sources of complex data.
- Employing intelligent ETL to cope with complex data integration issues.
- The absence of browser for parsing Active XML documents.
- Conceiving how to intelligently organize the ETL tasks.

Proposal approach features:

- Integrate complex data into repository of Active XML documents.
- Incorporate active rules into integration tasks realizing intelligent ETL.
- Exploit XML to unify, model and store complex data.
- Utilize Web services to overcome the distribution, heterogeneity and non-interoperability problems of complex data.
- Provide excellent data freshness by invoking embedded AXML Web services.

Publications

- Rashed Khalil, Wail Elkilani, Nabil Ismail, Mohie Hadhoud, "A Cost Efficient Location Management Technique for Mobile Users with Frequently Visited Locations ", *Proceedings of the 4th Annual Communication Networks and Services Research Conference (CNSR'06) - Moncton, New Brunswick, Canada*, p.p. 259 - 266.
- Rashed Khalil, Wail Elkilani, Nabil Ismail, Mohie Hadhoud, "A Cost Efficient Location Management Technique using Replicated Databas ", *INFOS2006,, Cairo, Egypt*, March 2006.
- Rashed SALEM, Jérôme DARMONT & Omar BOUSSAID, "Active XML data warehouses for intelligent, on-line decision support: State of Art", *Technical report, ERIC, Lyon2, May 2008*.
- R. Salem, "Complex Data Integration into an Active XML Repository", *The International ACM Conference on Management of Emergent Digital EcoSystems (MEDES 09) - Student workshop, Lyon, France, October 2009; ACM*

Anna STAVRIANOU

Current Position : ATER, Ph.D. Student
E-mail : Anna.Stavrianou@univ-lyon2.fr
Web site : <http://recherche.univ-lyon2.fr/eric/92-Anna-STAVRIANOU.html>
Birth Date : 23/02/1977
Arrival Date : 01/10/2005
Research supervisor : Jean-Hugues CHAUCHAT & Julien VELCIN



Research topics

The development of Web 2.0 has resulted in the generation of a vast amount of online discussions. Mining and extracting quality knowledge from online discussions is significant for the industrial and marketing sector, as well as for e-commerce applications. Discussions of this kind encapsulate people's interests and beliefs and hence, there is a great interest in acquiring and developing online discussion analysis tools.

In our work, we propose a graph-oriented model. The vertices of the graph represent postings. Each posting encapsulates information such as the content of the message, the author who has written it, the opinion polarity of the message and the time that the message was posted. The edges among the postings point out a "reply-to" relation.

The proposed model is accompanied by a number of measures which facilitate the discussion mining and the extraction of knowledge from it. The representation of an online discussion in the proposed way allows a user to "zoom" inside the discussion. A recommendation of messages is proposed to the user to enable a more efficient participation inside the discussion.

Our contributions are summarized in the following:

- 1) A novel model for analyzing online discussions.
- 2) We provide definitions of measures in order to facilitate the analysis of a discussion.
- 3) Recommendation of key messages.
- 4) A System Prototype which allows the interaction of a user with an online discussion.

Publications

A.Stavrianou, J. Velcin, J-H.Chauchat, "PROG: A Complementary Model to the Social Networks for Mining Forums", International Journal of Social Network Mining (*submitted*)

A. Stavrianou, J. Velcin, J. Chauchat, "Definition and Measures of an Opinion Model for Mining Forums", 2009 International Conference on Advances in Social Networks Analysis and Mining, Athens, Greece, 2009

A. Stavrianou, J. Velcin, J. Chauchat, "A combination of opinion mining and social network techniques for discussion analysis", Numéro spécial de RNTI « Fouille de Données d'Opinions », 2009

A. Stavrianou, J. Chauchat, J. Velcin, "A Content-Oriented Framework for Online Discussion Analysis", 2009 IEEE International Symposium on Mining and Web (MAW 09), IEEE Computer Society, UK, May 2009

A. Stavrianou, J. Velcin, J. Chauchat, "Un cadre pour la représentation et l'analyse de débats sur le Web", Atelier Veille Numérique (EGC-VN 09), Strasbourg, Janvier 2009.

A. Stavrianou, E. Bahri, N. Nicoloyannis, "Text Mining Issues and Noise Handling in Health Care Systems", 9th International Conference on System Science in Health Care, Lyon, France, September 2008

A. Stavrianou, J. Chauchat, "Opinion Mining Issues and Agreement Identification in Forum Texts", Atelier FOuille des Données d'OPinions (FODOP 08)), Fontainebleau, Mai 2008.

A. Stavrianou, P. Andritsos, N. Nicoloyannis, "Overview and Semantic Issues of Text Mining", SIGMOD Record, Vol. 36, No. 3, September 2007, 23-34.

Zhихua WEI

Current Position : PhD student
E-mail : zhихua.wei@hotmail.com
Wei@univ-lyon2.fr

Birth Date : 22/01/1979
Arrival Date : 01/12/2006
Research supervisor : Jean-Hugues Chauchat & Duoqian Miao



Research topics

Chinese Language Text Classification (TC). When facing the sea of information resources, the objects of TC are more complicated and diversity. More and more researchers regard that multi-label TC is more suited for many applications. My dissertation analyses the difficulties and problems in multi-label TC and Chinese text representation based on algorithms for single-label TC and multi-label TC. Aiming at high dimensionality in feature space, sparse distribution in text representation and poor performance of multi-label classifier, this dissertation brings forward corresponding algorithms from different angles.

Focusing on the problem of dimensionality “disaster” when Chinese texts are represented by using n-grams, two-step feature selection algorithm is constructed. The method combines filtering rare features within class and selecting discriminative features across classes. Moreover, the proper value of “n”, the strategy of feature weight and the independency among features are discussed based on variety of experiments.

In order to improve performance of multi-label classifier and degrade computing complexity, a new TC method based on LDA model is applied for Chinese text representation. It extracts topics statistically from texts and then texts are represented by using the topic vector. It shows competitive performance both in English and in Chinese corpus.

To enhance the performance of classifiers in multi-label TC, a compound classification framework is raised. It partitions the text space by computing the upper approximation and lower approximation. This decomposes a multi-label TC problem into several single-label TCs and several multi-label TCs which have less labels than original problem. That is, an unknown text should be classified by single-label classifier when it is partitioned into lower approximation space of some class. Otherwise, it should be classified by corresponding multi-label classifier.

An application system MLWC (multi-label web classifier) was designed. It could call the result from Search Engines directly and classify these results real-time using improved Naïve Bayes classifier. This makes the browse process more conveniently for users. Users could locate the texts interested immediately according to the class information given by MLWC.

Key Words: Chinese text, multi-label classification, rough set, LDA (Latent Dirichlet Allocation), text representation, classifier design

Publications

1. Z. Wei, D. Miao, J-H Chauchat and R. ZHAO, 2009, “N-grams based feature selection and text representation for Chinese Text Classification”, Special issue of the International Journal of Computational Intelligence Systems (IJCIS)
2. Z. Wei, D. Miao, J-H Chauchat and C. Zhong, 2008, “Feature Selection on Chinese Text Classification Using Character N-grams”, RSKT2008, The Third International Conference on Rough Sets and Knowledge Technology, China, Chongqing
3. Miao D., Wei Zhихua, (Book) 2007, “Chinese Language Understanding Algorithms and Applications” by Tsinghua University Press
4. D. Miao, Min Chen, Z. Wei, Qiguo Duan 2006, “A Reasonable Rough Approximation for Clustering Web Users”. The WICI International Workshop on Web Intelligence (WI) meets Brain Informatics (BI), WImBI: 428-442
5. Z. Wei, Miao D. et al. A 2006, New Structure-based Bill Location Technique, Computer Application.No.10.2006.

Ecole Doctorale
Interdisciplinaire Sciences-Santé
(EDISS)

Cynthia BASILEU

Current Position : PhD student
E-mail : cbasileu@yahoo.fr
Web site :
Birth Date : 29/04/1983
Arrival Date : 01/09/2009
Research supervisor : Ahmed Bounekkar,
Nadia Kabachi, Michel Lamure



Research topics

Sujet : Modélisation d'un système de gestion de crise sanitaire.

Résumé : Dans le but de gérer une situation de crise sanitaire lié à l'apparition d'une pandémie de grande envergure, le problème de contenir cette pandémie tout en préservant la capacité de production de l'économie nécessite d'être capable de prendre très rapidement les mesures adéquates. C'est pourquoi nous nous sommes focalisés sur le développement d'un modèle d'aide à la décision à base d'agents hybrides simulant la diffusion d'une pandémie, fondé sur les caractéristiques médicales de la pandémie ainsi que sur la structure socio-économique de la zone géographique concernée. Dans ce sens, une modélisation prétopologique est proposée. La structure socio-économique étant au cœur du modèle, une modélisation prétopologique du concept de réseau social est proposée et intégrée dans l'approche agent.

Mots-clés : modélisation, pandémie, systèmes dynamiques, gestion de crise, prétopologie, simulation, système multi-agents.

Publications

Conférence Internationale :

C. Basileu, A. Bounekkar, N. Kabachi, M. Lamure, J-M Pingeon, D. Roy. Modèle de diffusion spatiale de pandémie basé sur les flux migratoires quotidiens. Proceedings of the ninth International Conference on System Science in Health Care (ICSSHC 2008). Lyon, 3-5 septembre 2008.

Conférences Francophones :

C. Basileu, L. Lamarsalle, M. Lamure, A. Vainchtock. Analyse statistique de données de chaînage à partir du PMSI. Congrès CALASS'07, Marseille.

C. Basileu, S. Ben Amor, M. Bui, A. Bounekkar, N. Kabachi, M. Lamure et M. Toumi. Stochastic networks. 10ème journées Francophones "Extraction et Gestion des Connaissances" (EGC 2010). Hammamet, 26-29 janvier 2010.

Alain FALANGA

Current Position : PhD student
E-mail : a.falanga@ch-brignoles.fr
Web site :
Birth Date : 15/03/1971
Arrival Date : 01/01/2009



Research supervisor : HDR - Mr Stéphane Bonnevey

Research topics

La nouvelle réforme dans les hôpitaux a modifié en profondeur la manière dont l'hôpital fonctionne. Précédemment les hôpitaux fonctionnaient avec des dotations globales (enveloppes financières attribuées en fonction de différents critères, comme par exemple le nombre de lits ou le bassin de population). Ce mode est en train de disparaître au profit d'une dotation à l'activité (T2A), c'est l'activité produite qui définit les sommes qui sont allouées à chaque établissement. Dès lors le "pilotage" des établissements sur la manière dont les activités évoluent devient hautement stratégique. Le système d'information regorge d'informations qui peuvent permettre de mieux "comprendre" comment l'activité évolue (Données PMSI, de facturation, médicales).

L'objectif de ces travaux de recherche est :

- De fournir des outils pour la collecte et l'échange des flux d'informations au sein du SIH en se basant pour les échanges sur des normes comme le HL7 et l'XML, ainsi que des technologies d'interopérabilités et de partages d'informations éprouvées (Agents, MOM). L'objectif est de construire un système qui soit faiblement couplé (par la mise en place par exemple de dictionnaires métiers applicatifs et de description objets indépendante du système (Description XML)) avec le SIH et qui offre la capacité de fonctionner en "temps réel" et ceci indépendamment de la charge du système. L'utilisation de technologies de type Open Source (ActiveMQ, PostgreSQL, Python) doit permettre l'évolutivité du système.
- De modéliser des systèmes qui permettent de définir de manière prédictive la trajectoire dans patient dans le cadre de sa prise en charge au niveau d'un service d'urgence. L'objectif étant de pouvoir a priori et à partir des données connues aider à l'évaluation du temps de prise en charge du patient (en termes de temps).

Les travaux sont menés autour des référentiels, études et travaux produits par la MeaH, le GMSIH, l'Observatoire Régional des Urgences de PACA, et l'association HL7.

Mot clés : MOM, Broker, HL7, Messaging, Multi agent, Scripting, Trajectoire, Analyse, MeaH, ORU Paca, GMSIH, Open Source.

Publications

Salah HADJAB

Current Position : Etudiant en 1ère année de thèse de doctorat
E-mail : hadjabsalah@yahoo.fr
Web site : <http://salah.hadjab.free.fr/>
Birth Date : 29/04/1978
Arrival Date : 01/12/2009
Research supervisor : Directeur de thèse : Stéphane Bonnevey (EDISS)
Co-directeur de thèse : Nadia Kabachi (EDISS)



Sujet de recherche : « Modèles décisionnels pour la négociation entre agents rationnels »

Résumé : l'objectif est de travailler sur les problématiques de la négociation compétitives et coopératives dans les systèmes multi-agents. La négociation apparaît comme une technique de résolution de conflits. En effet, une situation de conflit peut être résolue par l'obtention d'un accord ou d'un consensus résultant d'un processus de négociation.

En général, la recherche sur la négociation peut être décomposée en trois larges thèmes :

- Les modèles et protocoles de négociation (génériques) : l'ensemble des règles qui gouvernent l'interaction ;
- Les objectifs de la négociation : l'intervalle des critères pour lesquels une solution doit être atteinte ;
- Les modèles décisionnels des agents : le moyen par lequel un agent atteint ses objectifs de négociation tout en suivant les règles de la négociation.

Le travail demandé se placera sur le thème des modèles décisionnels pour la négociation entre agents rationnels. Pour cela, une plateforme doit être développée pour pouvoir simuler et valider le travail de modélisation.

Ce travail de modélisation et de simulation sera appliqué à des problématiques de décision dans le secteur de la santé.

Publications

Rapport de DEA RACOR "Réseaux Avancés de Connaissance et Organisations" : Assemblage de composants logiciels : les composants composites. French Master's Thesis (DEA). July 2003. (In french).

Ahmed KAFIFY

Current Position : PhD student
E-mails : akafify@bat710.univ-lyon1.fr
AHMED.KAFIFY@bvra.etu.univ-lyon1.fr
Web site : -
Birth Date : 12/10/1980
Arrival Date : 01/2009
Research supervisor : Stéphane BONNEVAY and Ahmed BOUNEKKAR



Research topics:

Title:“Intelligent metaheuristic for multiobjective decision support-application on health care system”

Metaheuristics based techniques have attracted the attention of a lot of researchers in dealing with decision support and optimization problems, especially in combinatorial optimization NP-hard problems. Many real world decision making problems involve simultaneous optimization of multiple objectives. In this case, there exists a set of solutions which are superior to the rest of solutions in the search space when all objectives are considered but are inferior to other solutions in the objective space. These solutions are known as Pareto optimal set. Since none of these solutions is absolutely better than any other one. The choice of one solution over the other requires knowledge and preference information.

A metaheuristic is an iterative generation process which guides a subordinate heuristic by combining intelligently different concepts for exploring and exploiting the search space to find efficiently near-optimal solutions. Metaheuristics can be classified as: Population based versus single point search which called trajectory Methods, Memory versus memory-less methods...etc.

Recent research has shown that the hybridization of metaheuristics is a powerful mechanism to develop more robust and efficient methods to solve hard optimization problems. The combination of different techniques and concepts behind metaheuristic, if well designed has the potential to exploit their advantages while diminishing their drawbacks. The hybridization can be performed through the combination of two or more metaheuristics or metaheuristics with another kind corresponds to embedded module such as data mining (DM). Data mining refers to the extraction of new and potentially useful knowledge from datasets in terms of patterns and rules. It is responsible for identifying frequent patterns of good quality solutions. The basic concept of incorporating a Data Mining process in metaheuristic is that patterns found in high quality solutions obtained in earlier iterations can be used to conduct and improve the coming iterations which lead to better solutions than pure metaheuristic.

Our research tend to achieve the following aims:

- Improving the efficiency and the effectiveness of the existed methods by incorporating some useful techniques such as Data Mining.
- Study how to develop and implement new techniques using hybridization that have the ability to avoid the drawbacks and shortcoming of the pure metaheuristics and to enhance the performance and the quality of the solutions obtained.
- Study the influence of the improved and developed techniques in treating Multi-objective optimization problems especially NP-Hard combinatorial optimization problems with respect to Time consumption, quality of obtained solutions, etc.
- Application of the improved techniques on the health care system.

Publications:

A. Kafify, M. Sherbiny, W. Abd El-Wahed and N. Ismail, “A Hybrid Multiobjective Particle Swarm-Genetic Algorithm (PSO-GA) Approach With Pareto Archiving,” *in the proceeding of the 5th International Conference on Informatics and Systems (INFOS2007)*, Faculty of Computers and Information, Cairo University, Giza, Egypt, 24-26 March 2007.

A. Kafify, M. Sherbiny, W. Abd El-Wahed and N. Ismail, “Using Analytical Hierarchy Process (AHP) to select the most suitable efficient solution based on DM preferences ”, *in the proceeding. of the 5th International Conference on Informatics and Systems (INFOS2007)*, Faculty of Computers and Information, Cairo University, Giza, Egypt, 24-26 March 2007.

Franck MAUNOURY

Current Position : Phd student
E-mail : franck.maunoury@statesia.com
Web site : www.statesia.com
Birth Date : 09/07/1971
Arrival Date : 01.2007
Research supervisor : Jean-Paul AURAY



Thesis title is CAN THE FRENCH GENERAL PRACTITIONER AS A GATEKEEPER BE COST-EFFECTIVE FOR MANAGING CHRONIC PATIENTS TREATED WITH INHALED CORTICOSTEROIDS? The defense was performed in November, 5th 2009.

METHODOLOGY: My research is based on health economics methods as cost-effectiveness analysis. The French Regional Health Insurance database (named Erasme) has been analysed for assessing the cost-effectiveness impact of the French general practitioner considered as a gatekeeper since the French Health Insurance 2004 reform has established the principle of “Coordinated care pathway”. Multivariate data analysis techniques were used in order to try a response for the question research.

SUMMARY: The objective of this thesis is to conceptualize, starting from the exploitation of the refunding care data from the Sickness insurance fund, the various care pathways introduced by chronic diseases as asthma, and to study their determinants by analysing the profile and the subsequent behavior of the general practitioner. The study of the relation between the prescriptive behavior and the care trajectory is carried out by different multivariate analyses. The other objective is to assess, from a pharmacoeconomic point of view, the impact of the general practitioner characteristics on these various care pathways followed by the patients with chronic diseases. The characteristics likely to identify a typology of practitioners correspond to the variables influencing the prescriptive behavior (age, sex, duration of exercise, type of exercise, etc).

The principal question of the thesis is that of the regulating effect of the economic incentive, rested on the coordinated care pathway (reform of the Sickness insurance, August 2004), on the care pathways, really observed by the chronic patients. The corollaries are: Does the no-reimbursement of some medical acts, not considered in the coordinated care pathway, have a significant impact on the empirical recourse of the patient? Does the profile of the general practitioner have an effect on the respect or not of the allowed care trajectory classified by the French reform? What are the principal determinants of disregarding this referential trajectory, by notably analysing the “practitioner - patient” characteristics?

DISCIPLINE: Health economics

KEY WORDS: General practitioner, cost-effectiveness analysis, respiratory disorder, propensity score, semi-Markovian modeling.

Publications

- « Le processus de décision markovien : Adaptation des routines "msm" pour l'évaluation médico-économique des stratégies de santé ». 9th International conference on system science in health care. Septembre 2008.
- « Evaluation médico-économique des stratégies de santé : Application des processus de décision markoviens sous le logiciel R. ». Revue Santé Décision Management. Article vol. 11/3-4 - 2008 - pp.169-181 - doi:10.3166/sdm.11.3-4.169-181. Février 2009.
- Article accepté in Journal d'Economie Médicale : “Analyse coût efficacité du parcours de soins coordonnés des insuffisants respiratoires chroniques traités par corticostéroïdes inhalés” – Septembre 2009.
- “French gatekeeping cost-effectiveness impact on chronic patients treated with inhaled corticosteroids, in real life”. International Society for Pharmacoeconomics and Outcomes Research: 12th Annual European Congress « Health care decision making in Europe: From Patients to Populations”. Octobre 2009.
- “Can the French gatekeeping system be cost-effective for patients with chronic respiratory disorder?” International Society for Pharmacoeconomics and Outcomes Research: 12th Annual European Congress « Health care decision making in Europe: From Patients to Populations”. Octobre 2009.

Nathalie NICOLLIN épouse NOMDEDEU

Current Position : Manager
E-mail : nnomdedeu@mgen.fr
Web site : NC
Birth Date : 06/03/1964
Arrival Date : 12.2007



Research supervisor : Jean-Paul AURAY

Le titre de la thèse est ECONOMIE DES RESSOURCES HUMAINES EN ETABLISSEMENT DE SANTE.

Thesis title is HUMAN RESOURCES ECONOMY IN HEALTHCARE ORGANIZATIONS.

METHODOLOGIE : notre recherche se réalise en privilégiant une démarche abductive, inspirée de la « *Grounded Theory* »

METHODOLOGY: Our research is carried out through abductive approach, inspired by the "Grounded Theory".

RESUME : Les ressources humaines ont une place particulière dans les nouvelles théories économiques. L'élaboration de nouvelles politiques économiques et la construction de nouveaux tableaux de bord à partir d'indicateurs spécifiques basés sur le facteur humain est nécessaire pour réussir une nouvelle politique économique en santé au sein des établissements de soins impliquant les professionnels de santé.

SUMMARY: human resources have a special place in the new economic theories. Development of new economic policies and construction of new dashboards from specific indicators based on the human factor is necessary for successful economic reform policy in health with professionals of health.

DISCIPLINE : SCIENCE ECONOMIQUE

DISCIPLINE: economics

MOTS CLES : capital humain, ressources humaines, établissement de santé, management, compétences, connaissances, indicateurs RH, politique économique en santé.

KEY WORDS: human capital, human resources, health establishment, management, skills, knowledge, HR indicators, Economic policy in health.

Publications (5 maxi)

- Un idéal Ressources Humaines pour gérer un établissement de santé? Enjeux hospitaliers septembre 2006
- L'évaluation professionnelle des paramédicaux - Objectifs soins n°151 décembre 2006.
- Nouvelles technologies de l'information et gouvernance des systèmes de santé – 9^{ème} conférence sur la science des systèmes de santé – 5 septembre 2008

Florence RANCHON

Current Position : Doctorant
Pharmacien Assistant spécialiste

E-mail : ranchonflorence@yahoo.fr
florence.ranchon@chu-lyon.fr

Web site : aaaaa

Birth Date : 14/03/1981

Arrival Date : 2007



Research supervisor : Claude Dussart

Research topics

L'iatrogénie induite par les erreurs médicamenteuses est un problème majeur de santé publique, tant en termes de morbi-mortalité que de coût. Dans la littérature, les erreurs médicamenteuses en cancérologie sont largement décrites ; leurs conséquences cliniques pour les patients sont multiples : préjudice temporaire ou permanent, décès du patient. Pourtant, les facteurs de risque de survenue des erreurs médicamenteuses restent mal connus, retardant la mise en place d'actions préventives spécifiques concrètes. Des actions de détection, prévention et évaluation de ces erreurs médicamenteuses sont cependant indispensables afin de garantir la sécurité des patients, et sont d'autant plus attendues que la toxicité intrinsèque des molécules anticancéreuses, leur marge thérapeutique étroite et l'évolutivité de la pathologie potentialisent les conséquences cliniques de ces erreurs. Par ailleurs le coût induit par l'iatrogénie médicamenteuse est peu évalué. Très peu de données sont disponibles sur les conséquences économiques des événements indésirables des soins, particulièrement en ce qui concerne les erreurs médicamenteuses en cancérologie.

Cette thèse, réalisée au sein du Centre Hospitalier Lyon Sud (Hospices Civils de Lyon) s'inscrit dans un objectif d'amélioration de la qualité des soins et se distingue en trois axes.

1/ Le premier axe a pour but d'évaluer l'importance des erreurs médicamenteuses (erreurs avérées ou interceptées et qui n'ont donc pas atteint le patient) en terme de fréquence, nature, conséquences cliniques pour les patients et de coût d'opportunité pour l'assurance maladie.

2/ Le deuxième axe cherche à identifier des facteurs de risques de survenue de ces erreurs médicamenteuses par l'identification de patients et de séjours hospitaliers à risque. L'outil de prévention développé repose sur la modélisation de l'iatrogénie médicamenteuse en cancérologie et la détermination, à partir des différents facteurs de risque identifiés, d'un score de risque prédictif de la survenue d'une erreur pour un patient et un séjour donnés.

3/ Enfin, le dernier axe a pour objectif la mise en place continue d'actions concrètes et pratiques de prévention de l'iatrogénie médicamenteuse en cancérologie. La prévention repose sur des mesures individuelles (prise en charge spécifique pour « les patients et les séjours à risques » identifiés par le score de risque à l'entrée) et collectives (mise en place de Revue des Erreurs liées aux Médicaments dans les services de cancérologie).

Publications

- Cancer chemotherapeutic error in hospitalized patients: attributable damages and extra costs. Revue visée : Journal of Clinical Oncology
- 2009 : Thèse pour l'obtention du Diplôme d'Etat de Pharmacie Hospitalière et des collectivités. « Impact médico-économique de la prévention des erreurs en cancérologie : étude prospective de 1 an »
- Communication orale: Medico-economic evaluation of chemotherapy errors prevention: a prospective study. Ranchon F et al. 37th European symposium on clinical pharmacy, 22-24 Octobre 2008, Dubrovnik, Croatie.

Thibaut Sanglier

Current Position : PhD student
E-mail : thibautsanglier@gmail.com
Web site :
Birth Date : 22/10/1980
Arrival Date : 06/02/2007



Research supervisor : Jean Paul Auray

Research topics

Major depressive disorder in the elderly is increasingly prevalent and is expected to be the 2nd most frequent cause of disability by 2020. Antidepressants have been found effective in the elderly. Because elderly population is associated with multiple chronic morbidities and polypharmacy expected course of antidepressant treatment maybe disturbed. That is why treating MDD in the elderly raise several challenges when compared with younger adult population.

Randomized controlled trials provide an accurate assessment of treatment efficacy. Randomized controlled trials methodology relies on a limited and selected population and command very specific and careful patient follow-up. Although efficacy can be demonstrated effects and uses of a compound in general population may lead to unexpected adverse events, patterns of use or effectiveness.

Epidemiology aims to provide real life observation by describing and assessing patients in routine practice. Data is collected without any kind of intervention on the provided care.

Healthcare administrations created de-identified large database collecting healthcare expenditures of their insured patients. This data is known to be very reliable regarding drug exposure and dispensation. Observing and assessing patient profiles, drug exposure dynamics and medical expenditure of patients provide a unique opportunity to assess treatment patterns of prescription and real life outcomes.

Few studies have focused on the comparison of the treatment provided to the elderly with the younger adult population using large claim databases. Highlighting specificities of antidepressant treatment provided to depressed elderly population in a real life setting would help to identify specific unmet needs and understand how differently the depression is managed in this specific population.

Publications

Sanglier T, Saragoussi D, Milea D. **A comparison of persistence and healthcare costs related to different treatment strategies after initial escitalopram 10mg in major depressive disorder.** Presented at ISPOR 11th annual european congress, Athens, November 2008.

Sanglier T, Saragoussi D, Milea D, Auray JP, Tournier M. **Characteristics of depressed elderly patients compared with younger depressed adult patients: an assessment based on a large U.S. insurance claims database.** Presented at ISPE US, Providence, August 2009.

Sanglier T, Saragoussi D, Milea D, Auray JP, Tournier M. **Treatment patterns in depressed elderly and younger adult patients in the U.S..** Presented at ISPE US, Providence, August 2009.

Sanglier T, Saragoussi D, Milea D. **Comparison of different treatment options following initiation dose of 10mg escitalopram.** Presented at American Psychiatric Association U.S., San Francisco, May 2009.