

## List of Accepted Papers for Track ASAI2009

**Asistencia personalizada a grupos de usuarios: análisis y comparación de técnicas.** Ingrid Christensen and Silvia Schiaffino.

**Abstract:** Los sistemas de recomendación se utilizan para realizar recomendaciones de ítems potencialmente interesantes para un usuario en variados dominios. Existe un gran número de dominios que sugieren la necesidad de proveer técnicas de personalización para grupos de usuarios y no sólo focalizarse en usuarios individuales. En este trabajo se presentan dos aplicaciones que implementan técnicas de generación de recomendaciones grupales a partir del cual se realizaron dos instanciaciones en dominios diferentes: recomendación de música y recomendación de películas. A partir de estas aplicaciones se analizaron y compararon las distintas técnicas evaluando su funcionamiento en distintos dominios de aplicación.

**Feature selection on wide multiclass problems using OVA-RFE.** Pablo M. Granitto and Andrés Burgos.

**Abstract:** Feature selection is a pre-processing technique commonly used with high-dimensional datasets. It is aimed at reducing the dimensionality of the input space, discarding useless or redundant variables, in order to increase the performance and interpretability of models. For multiclass classification problems, recent works suggested that decomposing the multiclass problem in a set of binary ones, and doing the feature selection on the binary problems could be a sound strategy. In this work we combined the well-known Recursive Feature Elimination (RFE) algorithm with the simple One-Vs-All (OVA) technique for multiclass problems, to produce the new OVA-RFE selection method. We evaluated OVA-RFE using wide datasets from genomic and mass-spectrometry analysis, and several classifiers. In particular, we compared the new method with the traditional RFE (applied to a direct multiclass classifier) in terms of accuracy and stability. Our results show that OVA-RFE is no better than the traditional method, which is in opposition to previous results on similar methods. The opposite results are related to a different interpretation of the real number of variables in use by both methods.

**Gaining knowledge of data structure using stability concepts.** Ariel Baya and Pablo Granitto.

**Abstract:** In this work we propose the use of stability concepts to gain knowledge on data structure. To analyze stability we use an exploration algorithm based on the idea of "Sub-sample and Cluster". In the clustering step we use PAM and HC-av to construct solution ensembles and then we evaluate their similarity with the Variation of Information Criterion. We apply our method to analyze artificial and real data. We use some datasets to compare our method with another analysis technique proposed by Ben-Hur and Guyon. We show that we can obtain comparable information with both methods but our new technique presents results that can be more easily analyzed.

**A Multi-Objective Evolutionary Algorithm Approach to Learn Disjunctive and Conjunctive Topical Queries.** Rocío Luján Cecchini, Carlos Martín Lorenzetti and Ana Gabriela Maguitman.

**Abstract:** Topical search refers to the process of formulating queries that reflect a

thematic context. A combination of machine learning and information retrieval techniques can be applied to automate this process. In this work we propose to apply single- and multi-objective evolutionary algorithms to automatically evolve a population of topical queries. We report on the results of different strategies that attempt to evolve conjunctive and disjunctive queries. In our evaluations we observe that disjunctive queries have the potential to achieve better retrieval performance than conjunctive queries. In addition, our analysis reveals the limitations of the single-objective approach and highlights the advantages of applying multi-objective evolutionary algorithms for the problem at hand.

**A Recommender System for Classification of Non-Functional Requirements.**  
Agustin Casamayor, Daniela Godoy and Marcelo Campo.

**Abstract:** The early detection and classification of non-functional requirements (NFRs) is not only a hard and time consuming process, but also crucial in the evaluation of architectural alternatives starting from initial design decisions. In this paper, we propose a recommender system based on a semi-supervised learning approach for assisting analysts in the detection and classification of NFRs from textual requirements descriptions. Classification relies on a reduced number of categorized requirements and takes advantage of the knowledge provided by uncategorized ones as well as certain properties of text. Experimental results show that the proposed recommendation approach based on semi-supervised learning outperforms previous proposals for classifying different types of requirements.

**Obtención de Comportamientos Emergentes en Sistemas Multiagente mediante Aprendizaje por Refuerzo.** Cristina Parpagione and Juan Miguel Santos.

**Abstract:** Este trabajo direcciona el problema de encontrar comportamientos individuales a partir de un comportamiento grupal objetivo en un sistema multiagente usando aprendizaje por refuerzo. Lo que se espera es que el grupo de agentes, usando sus comportamientos individuales den lugar a un comportamiento emergente coincidente con el comportamiento objetivo. Como prueba de concepto, se aplicó el método al problema de agrupamiento. El aprendizaje se llevó a cabo en un grupo reducido de agentes y luego se validó de dos formas: 1) resolviendo el problema a partir de los comportamientos obtenidos con los pocos agentes que habían realizado el aprendizaje y; 2) dotando a todos los agentes, incluso los no entrenados, con los comportamientos obtenidos y comprobando así, generalización. Los agentes utilizados son simples, con una acotada capacidad de sensado y con muy baja comunicación entre ellos.

**Q-Function Kernel Smoother: a New Approach for Opened Issues in Huge State-Action Spaces.** Ana Julia Villar and Juan Miguel Santos.

**Abstract:** Reinforcement learning is a method where agents learn how to map states to actions. They have to interact with their environment and then, they receive reinforcements from it. Q-learning is, probably, the most used technique in reinforcement learning and it needs to compute by successive estimations an action-value function (Q-function). In a large number of problems, the environments have huge state-action spaces and, therefore it is necessary to compute Q for countless pairs of states and actions. A previous work has shown a way to do that using only one approximator of Q, a kernel smoother, over the state-action space, where resolution depends on the frequency of visits during the learning. This paper extends the previous proposal presenting improvements on some issues related to its use. To validate the

proposed ideas, experiments by simulations were performed and their results are presented.

**Towards matrix-based selection of feature pairs for efficient ADMET prediction.**

Marc Strickert, Axel Soto, Jens Keilwagen and Gustavo Vazquez.

**Abstract:** Time-consuming and expensive experimental assessment of physicochemical compound properties can be alleviated by computational intelligence methods if a specific classification or regression task is connected to them. In this case the properties relevant for such task can be determined by feature selection methods. Due to the combinatorial explosion, strict feature selection is a computationally expensive task. Yet, feature weighting can be done by rescaling synchronously all axes (features) of the data space, for example, in a way that optimizes class separation. Here, a recently proposed adaptive rank-1 matrix distance resembling the Mahalanobis distance is used to assess the weight of pairs of compound features related to the magnitude of the octanol-water partitioning coefficient. From top-rated feature pairs, regression models based on neural networks were successfully generated.

**An Argumentative Intentional Model for High Level Reasoning of Mobile Robots.**

Sebastian Gottifredi, Mariano Tucut, Alejandro Garcia and Guillermo R. Simari.

**Abstract:** In this work we present an argumentation based intentional model for a BDI architecture defined formally for high level control of mobile robots. The proposed approach provides a sophisticated way of handling conflicting intentions using the argumentative formalism of Defeasible Logic Programming. To manage the intentional model, we present a special notion of argument disagreement and a new comparison criterion. Finally an implementation for the proposed system is introduced.

**A Proposal for Simplifying Explanations from Ensembles of Symbolic Classifiers.**

Flavia Bernardini, Maria Carolina Monard and Ronaldo Prati.

**Abstract:** Data mining applications generally use learning algorithms in order to induce knowledge. To accomplish this task, these algorithms should be able to operate with massive data sets. Several techniques, such as data sampling, can be used to scale up learning algorithms to deal with large datasets. Using data sampling, learning algorithms can be applied to small samples of the original dataset, and the individual classifiers can then be combined into an ensemble which, in numerous situation, can be more accurate than the individual classifiers. However, ensembles often lack the facility to explain their decisions. This work explores a method to offer a concise explanation of ensembles decisions whenever the ensembles are composed by a combination of symbolic classifiers. Different methods used to construct ensembles are also described.

**Minimum Classification Error Training of Hidden Markov Models for Sequential Data in the Wavelet Domain.** Diego Tomassi, Diego Milone and Liliana Forzani.

**Abstract:** In the last years there has been increasing interest in developing discriminative training methods for hidden Markov models, with the aim to improve their performance in classification and pattern recognition tasks. Although several advances have been made in this area, they have been targeted almost exclusively to standard models whose conditional observations are given by a Gaussian mixture density. In parallel with this development, a special kind of hidden Markov models defined in the wavelet domain has found wide-spread use in the signal and image processing community. Nevertheless, these models have been typically restricted to fully-tied parameter training using a single sequence and maximum likelihood estimates. This paper takes a step forward in the development of sequential pattern recognizers based on wavelet-domain hidden Markov models by introducing a new discriminative training method. The learning strategy relies on the minimum classification error approach and provides reestimation formulas for fully non-tied models. Numerical experiments on a simple phoneme recognition task show important improvement over the recognition rate achieved by the same models trained under the maximum likelihood estimation approach.

**Evolutionary Artificial Immune System Optimization.** Andre Barreira, Carlos Eduardo Oliveira, Otavio Teixeira and Roberto Oliveira.

**Abstract:** This article presents a new evolutionary algorithm of advanced optimization based in the technique of Artificial Immune Systems, more specifically, in the inclusion of evolutionary characteristics in the algorithm AISO. For this relative aspects to the subjects of theoretical recital are presented, the algorithm in itself and finally some simulation and comparison of the algorithms with one real world problem.

**Markerless 3D Head Pose and Facial Expression Tracking.** Lucas Daniel Terissi and Juan Carlos Gómez.

**Abstract:** In this paper, novel markerless 3D head pose and facial expression tracking algorithms, based on monocular image sequences (single camera), are presented. The proposed method is based on a combination of feature-based and model-based approaches for pose estimation. A generic 3D face model, which can be adapted to any person, is used for the tracking. In contrast to other methods in the literature, the proposed method does not require a training stage. It only requires an image of the person's face to be tracked facing to which the model is fitted manually through a graphical user interface. To evaluate the performance of the algorithms a video database was compiled. Simulation results show that the proposed tracking algorithms correctly estimate the head pose and facial expression, even when occlusions, changes in the distance to the camera and presence of other persons in the scene, occur. Although the implementation of the algorithms was not optimized for speed, they run near real time.

**An autonomous labeling approach to SVM algorithms for network traffic anomaly detection.** Carlos Catatania, Facundo Bromberg and Carlos García Garino.

**Abstract:** In the past years, several support vector machines anomaly detection approaches have been proposed in the network intrusion detection field. The main advantage of these approaches is that they can characterize normal traffic when trained using a data set containing not only normal traffic but also possible attacks. Unfortunately, these algorithms seem to be accurate only when the normal traffic vastly outnumbers the numbers of attacks or anomalies present in the dataset.

In this work we present an approach for autonomous labeling of normal traffic as a way of dealing with situations where class distribution do not present the required unbalance. The autonomous labeling process is made by SNORT, a misuse-based intrusion detection system. Experiments conduced on the 1998

DARPA dataset show the proposed autonomous labeling approach not only outperforms existing SVM alternatives but also obtains significant improvement over SNORT itself.

**Métodos de agrupamiento no supervisado para la integración de datos genómicos y metabólicos de múltiples líneas de introgresión.** Diego Milone, Georgina Stegmayer, Matías Gerard, Laura Kamenetsky, Mariana López and Fernando Carrari.

**Abstract:** Las numerosas aplicaciones de la inteligencia artificial a la biología de sistemas han dado lugar a nuevos algoritmos, además de la adaptación y reutilización de los existentes. En tareas de minería de datos se han aplicado diversos métodos estándar, como por ejemplo el bien conocido k-medias. Sin embargo, las capacidades de estos métodos es limitada en relación a otros algoritmos más recientes, tanto en su desempeño para el agrupamiento de patrones como para la representación e interpretación de los resultados obtenidos. En este trabajo se compara el desempeño de tres métodos de agrupamiento no supervisado en la tarea de integración y descubrimiento de relaciones entre variaciones en los contenidos de metabolitos y la expresion de genes de frutos de tomate. Los métodos considerados son el k-medias, el agrupamiento jerárquico y un método recientemente propuesto que se basa en mapas auto-organizativos. Se presentan los resultados obtenidos del análisis objetivo de la calidad de los agrupamientos y su significancia biológica. El modelo auto-organizado ha mostrado las más altas tasas de desempeño en las medidas de cohesión y separación, brindando además la máxima coherencia de las agrupaciones obtenidas desde el punto de vista del significado biológico.

**Array of Multilayer Perceptrons with No-class Resampling Training for Face Recognition.** David Capello, Cesar Martinez, Diego Milone and Georgina Stegmayer.

**Abstract:** A face recognition (FR) problem involves the face detection, representation and classification steps. Once a face is located in an image, it has to be represented through a feature extraction process, for later performing a proper face classification task. The most widely used approach for feature extraction is the eigenfaces method, where an eigenspace is established from the image training samples using principal components analysis. In the classification phase, an input face is projected to the obtained eigenspace and classified by an appropriate classifier. Neural network classifiers based on multilayer perceptron models have proven to be well suited to this task. This paper presents an array of multilayer perceptron neural networks trained with a novel no-class resampling strategy with takes into account the balance problem between class and no-class examples and increase the generalization capabilities. The proposed model is compared against a classical multilayer perceptron classifier for face recognition over the AT&T database of faces. The results obtained show interesting results regarding the improvement in classification rates.

**How k-Nearest Neighbor Parameters Affect its Performance.** Gustavo Batista and Diego Silva.

**Abstract:** The k-Nearest Neighbor is one of the simplest Machine Learning algorithms. Besides its simplicity, k-Nearest Neighbor is a widely used technique, being

successfully applied in a large number of domains. In k-Nearest Neighbor, a database is searched for the most similar elements to a given query element, with similarity defined by a distance function. In this work, we are most interested in the application of k-Nearest Neighbor as a classification algorithm, i.e., each database element has a label (class) associated, and the main goal of the algorithm is to decide the class of a new case based on the classes of the k most similar database elements. This work provides a discussion and presents empirical evidence of how the main parameters of k-Nearest Neighbor influence its performance. The parameters investigated are the number of nearest neighbors, distance function and weighting function. The most popular parameters choices were evaluated, including nine values for k, three popular distance measures and three well-known weighting functions. Our experiments were performed over third-one benchmark and "real-world" data sets. We recommend the use of the inverse weighting function and k=5 for HEOM and HMOM distance functions and k=11 to HVDM distance function.

**Agentes y Redes Bayesianas para Entrenamiento de Habilidades de Colaboración.**  
Rosanna Costaguta and Analia Amandi.

**Abstract:** Los sistemas de Aprendizaje Colaborativo Soportado por Computadora (ACSC) poseen reconocidas ventajas. Sin embargo, utilizar estos sistemas no garantiza un aprendizaje colaborativo efectivo. El éxito o fracaso de la experiencia de aprendizaje está condicionado por las habilidades colaborativas que manifiesten los estudiantes del grupo. En este artículo se presenta un modelo multiagente, aplicado al ámbito del ACSC, dedicado tanto al reconocimiento de conflictos en la dinámica grupal, como al entrenamiento personalizado de las habilidades de colaboración manifestadas por los estudiantes miembros de un grupo. El reconocimiento de conflictos se realiza aplicando el método de análisis de interacciones Interaction Process Analysis. La personalización se logra mediante redes bayesianas que consideran las características de colaboración del estudiante para aprender cual es la estrategia de entrenamiento más adecuada. El modelo fue implementado sobre un entorno de aprendizaje a distancia mostrando un alto grado de eficacia.

**Técnicas de NLP y WSD Asistiendo Al Desarrollo de Software Orientado a Aspectos.** Alejandro Rago and Claudia Marcos.

**Abstract:** El Desarrollo de Software Orientado a Aspectos (DSOA) provee medios sistemáticos para la identificación, modularización, representación y composición de crosscutting concerns en unidades denominadas aspectos. La ingeniería de requerimientos orientada a aspectos intenta identificar los posibles aspectos desde las primeras etapas del ciclo de vida de un sistema. La identificación de los aspectos en esta etapa mejora la trazabilidad entre los requerimientos y artefactos posteriores, facilita una estimación más sencilla del impacto del cambio y, particularmente, reduce el peligro de cambios no esperados en los productos de software. En este trabajo se presenta un enfoque de identificación de aspectos tempranos el cual realiza un análisis sintáctico y semántico de la funcionalidad del sistema especificada por medio de casos de uso. Para identificar los aspectos se utilizan técnicas de NLP y algoritmos WSD.