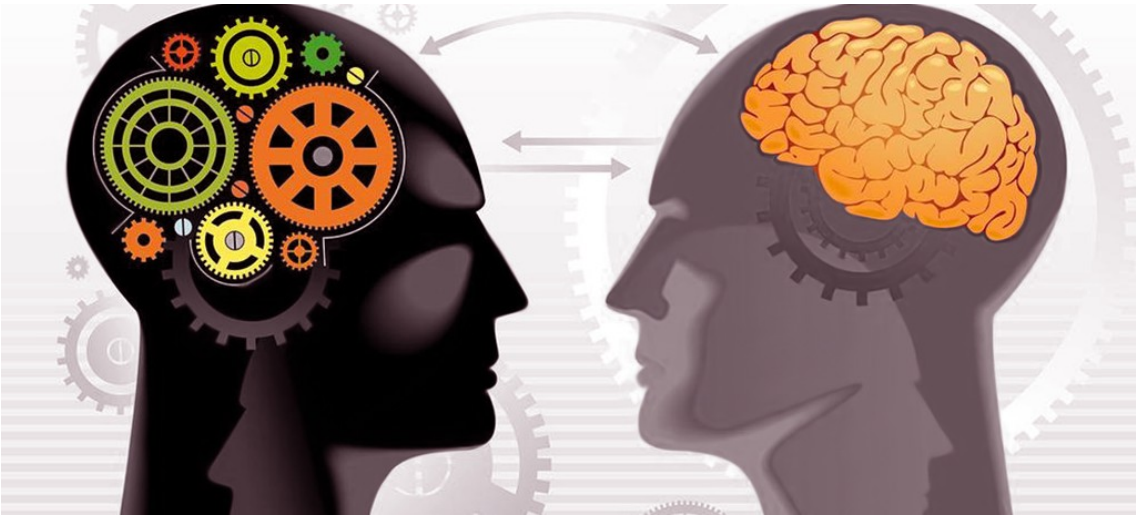
Programa de Doctorado de Sistemas Inteligentes, UNED

Doctoral Consortium /Jornadas de Doctorado

11-12 Junio/ June 11th-12th

Calle de Juan del Rosal 28040 Madrid

Cómo llegar: https://goo.gl/maps/er7Y3QQJYR92



Links for streaming

- Facultad de Educación, Salón de Actos: Enlace
- Facultad de Educación, Salón de Grados: Enlace
- ETSI Informática, Sala José Mira: Enlace

Panel 1: Diagnosis, Planning and control, Robotics and Artificial Vision

Experts: Dr. Marek J. Druzdzel (University of Pittsburgh)/Alfredo Cuesta (Universidad Juan Carlos I)
Sala J. Mira , ETSI Informatica, UNED , 4 planta

- Federico Castejón Lapeyra (2nd presentation, PT)
- Héctor E. Delgado-Ureña Poirier (2nd presentation, PT)
- Ricardo Navares Echegaray (2nd presentation, FT)
- Mikel Val Calvo (1st presentation, FT)

Panel 2: Intelligent access to multiligual and multimedia information & Teaching/Learning: collaboration and adaptation

Experts: Stefano Mizzaro (University of Udine), Edmundo Tovar (Universidad Politécnica de Madrid)
Sala de Grados, Facultad de Educación, UNED.

- Mario Almagro Cádiz (1st presentation, FT)
- Hermenegildo Fabregat Marcos (1st presentation, FT)
- Jennifer Hernández Becares (1st presentation, FT)
- Javier Rodríguez Vidal (2nd presentation, FT)
- Santa Vallejo Figueroa (2nd presentation, PT)

FT- full time
PT- part time

Panel 1: Diagnosis, Planning and control, Robotics and Artificial Vision			
José Mira Room (4th floor, UNED School of Computer Science)			
Monday June 11th		Tuesday June 12th	
10:00-11:15	Plenary talk:	10:00-11:15	Plenary talk:
11:15-11:40	Break	11:15-13:30	Panel evaluation meeting
11:40-12:30	F. Castejón		
12:30-13:20	H.E. Delgado-Ureña		
13:20-15:30	Lunch Break		
15:30-16:20	R. Navares		
16:20-17:00	M. Val		
17:00-17:40	S. Pérez Vasseur		
17:40-18:30	Panel evaluation meeting		

Panel 2: Intelligent access to multiligual and multimedia information & Teaching/Learning: collaboration and adaptation			
Salón de Grados (1st floor, UNED Faculty of Education)			
Monday June 11th		Tuesday June 12th	
10:00-11:15	Plenary talk:	10:00-11:15	Plenary talk:
11:15-11:40	Break	11:15-13:30	Panel evaluation meeting
11:40-12:30	J. Rodríguez		
12:30-13:20	S. Vallejo		
13:20-15:30	Lunch Break		
15:30-16:10	M. Almagro		
16:10-16:50	H. Fabregat		
16:50-17:30	J. Hernández		
17:30-18:30	Panel evaluation meeting		

Plenary talks:

10:00 AM Monday, June 11th
Prof. Marek Druzdzel
Title: Self-Confidence in Decision Support Systems Based on Bayesian Networks
Abstract: In this talk, I will focus on the problem of self-confidence of a decision support systems based on Bayesian networks. Self-confidence of such a system can be measured along three dimensions: (1) adequacy of the system's model to the case at hand, (2) the amount of information about the case at hand, and (3) the amount of knowledge used to construct the model, i.e., either the amount of experience of experts from whom the model originates or the amount of data available to train it. I will introduce the three measures and show how they can be operationalized in practical systems. There are several important applications of these measures, for example in recognizing cases that are likely to yield erroneous results. This is joint work with Marcin Kozniewski.

10:00 AM Tuesday, June 12th
Prof. Stefano Mizzaro
Title: TBD
Abstract: Effectiveness evaluation by means of a test collection is a standard methodology in information retrieval, with a long history. To gather relevance labels, the classical approach used in TREC-like initiatives was to use binary relevance judgments expressed by trained assessors. Two more recent trends are to rely on workers from the crowd as assessors, and to adopt multi-level relevance judgments, as well as gain-based metrics leveraging such multi-level judgment scales. After a brief introduction to test collection based evaluation, I will report on two experiments focusing on such fine-grained relevance scales. In some recent work (ACM SIGIR 2015, ACM TOIS 2017) we proposed unbounded relevance scales by means of magnitude estimation and compared them with multi-level scales. While magnitude estimation brings advantages like the ability for assessors to always judge the next document as having higher or lower relevance than any of the documents they have judged so far, it also comes with some drawbacks. For example, it is not a natural approach for untrained assessors to judge items as they are used to do on the Web (e.g., 5-star rating). In another more recent work (ACM SIGIR 2018) we proposed to collect relevance judgments over a 100-level relevance scale, a bounded and fine-grained scale having many of the advantages and dealing with some of the issues of magnitude estimation. The two approaches have been experimentally evaluated by means of large-scale crowdsourcing experiments, that compare the two scales with other traditional relevance scales (binary, 4-level). The results show the benefits of fine-grained scales over coarse-grained ones. Joint work with Shane Culpepper, Gianluca Demartini, Eddy Maddalena, Kevin Roitero, Mark Sanderson, Falk Scholer, Andrew Turpin.

Abstracts

Panel 1: Diagnosis, Planning and control, Robotics and Artificial Vision

1 NAME: FEDERICO CASTEJÓN LAPEYRA
THESIS DIRECTOR: ENRIQUE J. CARMONA SUAREZ
TITLE: AUTOMATIC DESIGN OF ANALOG ELECTRONIC CIRCUITS BY MEANS OF EVOLUTIONARY ALGORITHMS

ABSTRACT:
In the first part of this thesis, an algorithm for automatic synthesis of analog electronic circuits based on Grammatical Evolution (GE) is presented. GE is an evolutionary algorithm which can generate code in any programming language and uses variable length binary strings. The binary genome determines which production rules in a BNF grammar are used in a genotype to phenotype mapping process to a program. A suitable grammar for analog electronic circuit design has been developed for this task. Additionally, a specific crossover operator has been designed to avoid a destructive behavior of a standard crossover operator. Chromosome decoding generates a netlist which is evaluated using NGSpice. This method has been applied to seven case studies, and a comparison with previous works in the field based on Genetic Programming and other algorithms is shown. These results have been presented in a paper published in a scientific journal.
The second part of this thesis has focused on improving the success rate (SR) of the algorithm above mentioned. Several techniques have been tested to achieve this goal. First, a new technique for GE is presented. It is denoted by multigrammatical evolution and uses several grammars to decode an individual. An important SR improvement has been achieved by separating the two circuit synthesis stages (topology selection and circuit sizing) into two separate parts of the chromosome, and using two different grammars for the decoding process. Additionally, a suitable set of grammars and a specific crossover operator have been developed for this new algorithm. Third, parameter evolving has been tested to learn the maximum number of nodes for a circuit. For this, a syntax extension for the grammar has been used to implement parameter evolution and good results have been obtained with this technique. Finally, because the multigrammatical evolution algorithm tends to increment the number of components in the synthesized circuits compared to those circuits obtained with the seminal algorithm, a simple parsimony technique has been applied and analyzed.

2 NAME: HÉCTOR E. DELGADO
TITLE: IMPROVING THE COSMIC DISTANCE LADDER THROUGH HIERARCHICAL BAYESIAN SHRINKAGE OF GAIA PARALLAXES AND PHYSICAL PROPERTIES OF RR LYRAE STARS
PHD SUPERVISOR: LUIS M. SARRO

Abstract: RR Lyrae stars are primary standards candles that allow astronomers to infer extragalactic distances up to 100 kpc (kiloparsecs) by means of Period-Magnitude-Metallicity (PMZ) relationships. ESA's Gaia space mission DR1 and DR2 catalogues are providing accurate (but not exempt from uncertainty and systematic effects) trigonometric parallaxes (inverse of distances) for a set of RR Lyrae stars from the disk and halo of our Galaxy. We present a novel Bayesian hierarchical methodology that improves the calibration of RR Lyrae stars PLZ relationships using Gaia parallaxes, set the basis for a reassessment of the cosmic distance ladder and open new frontiers for the study of the morphology of objects from the Milky Way Local Group. The methodology uses the formalism of Bayesian networks to partition the parameter space associated to the problem of a calibrating PMZ relationship into several levels of statistical variability and take into account additional relationships between parameters that correct for biases and shrink the posterior estimates of parameters of interests.

3 NAME: RICARDO NAVARES
TITLE: FORECASTING URBAN AIR QUALITY THROUGH LSTM
PHD SUPERVISOR: JOSÉ LUIS AZNARTE

Abstract:

Knowing future pollution levels is of great interest for a wide range of research fields: from clinical institutions to regional representatives in charge of applying managerial environmental policies or even particular users who suffer from allergic and respiratory diseases. During the last decades, air quality has been gaining importance due to the health threats produced by high levels of environmental pollution. Within the context of this study, pollution is understood as both chemical pollutants and biotic factors present in the environment.

Despite the extensive literature and resources about the topic, very limited number of them take into account simultaneously both types of pollutants. Predicting chemical pollutants and airborne pollen concentrations based on meteorological conditions are two inherently different problems: atmospheric pollen concentrations depend on plant development during previous seasons which, in turn, depends on the climatological conditions during plant evolution. This implies long and mid-term relations between past atmospheric conditions and current plant status. Contrarily, chemical air pollutant levels are related to recent past atmospheric conditions. These differences suggest different approaches when predicting each problem. Also, in addition to the temporal dimension, it is important to take into account the spatial interactions between observation stations as they are implicitly related. In this work, we will present an unified approach based on deep learning Long Short Term Memory units which is able to overcome some of the limitations of traditional approaches. The system is able to capture the interactions and extract the relevant information,both temporal (long and short-term) and spatial, with a limited dependency on external expert intervention. The output of the system provides future environmental scenarios which will be applied, in the context of the thesis, in further developments to forecast hospital admissions.

4 NAME: MIKEL VAL
TITLE: EMOTIONAL HUMAN ROBOT INTERACTION USING A BCI INTERFACE.
PHD SUPERVISORS: JOSÉ RAMÓN ÁLVAREZ SÁNCHEZ, JOSÉ MANUEL FERRÁNDIZ

Abstract: The field of Emotional Robots begins to be perceived by many, as a tangible reality that seems to develop robustly in the field of research, promoting interdisciplinary approaches. Emotional or socially intelligent robots try to emulate human social intelligence, in order to be integrated into our society in a natural way. For this objective, the robots have to be able to learn both, to detect and reproduce the way humans communicate, given that in our personal relationships we use emotional signs that are expressed in non-verbal communication patterns, differences in the tone of voice or nuances in the semantics of the message used. That is why the main objective of this thesis is to improve the Robot-Human Interaction(HRI), to produce a more natural interaction for the user, under the hypothesis that the robot can improve its interaction if it is able to reduce its own uncertainty when detecting user emotions states. We will therefore investigate how to assess the emotional state of the user by adding information related to electrophysiological signals (pulse, skin conductance, temperature, EEG, etc ...), to the information that the robot is autonomously able to obtain by means of the analysis of the tone of voice, eye contact detection and facial gestures.

5 NAME: SEBASTIÁN PÉREZ VASSEUR
TITLE: Probabilistic Forecast of NO_2 levels in Madrid
PHD SUPERVISORS: JOSÉ LUIS AZNARTE

Abstract:
Pollution is becoming a main problem in cities where the high concentrations are becoming a health hazard and requiring urgent actions from the authorities (for example, cutting traffic). It is then required in some cases to be able to predict peaks of pollution in advance. With this objective in mind, we are studying and comparing different probabilistic forecasting techniques.

Panel 2: Intelligent access to multiligual and multimedia information & Teaching/Learning: collaboration and adaptation

1 NAME: JAVIER RODRÍGUEZ VIDAL
TITLE: AUTOMATIC GENERATION OF ONLINE REPUTATION REPORTS
PHD SUPERVISORS: JULIO ANTONIO GONZALO ARROYO, LAURA PLAZA MORALES

Abstract: Our long-term goal is the automatic generation of online reputation reports for organizations, brands and individuals using online (social) media, with a specific focus on Twitter (as the main source of fresh information). The third year has been devoted to two different tasks: 1) the detection and categorization of Twitter opinion makers, i.e. users whose opinions may influence the state of public opinion for some domain and 2) the automatic generation of extractive summaries. For the detection and categorization of opinion makers, we explore the use of the followers' posts as a novel feature. For the summarization task, we have used neural networks to build a vectorial representation based on word embeddings and employed this representation to compute the distance of the different tweets to the centroid of the different topics. This distance is used to rank the tweets and those that are more central to the topics are selected for the summary. The redundancy is removed using overlap between texts. Also, we have experimented with the authority and domain information extracted from the first task as a feature for summarization. Both tasks have been evaluated on different RepLab evaluation collections.

2 NAME: SANTA VALLEJO FIGUEROA

PHD SUPERVISOR: MIGUEL RODRÍGUEZ ARTACHO

TITLE: OPEN LINKED DATA TO FACILITATE THE MASHUP OF EDUCATIONAL RESOURCES

Abstract: Open Educational Resources (OERs) are very useful means for facilitating teaching and learning tasks, but its creation poses very challenges not only from Instructional point of view, the integration of tools and standards for its creation is a very hard and time-consuming task. Many researchers and institutions are interested on the generation, distribution and use of OERs. Regardless the instructional requirements, the integration of learning contents is a tendency and a challenge to promote the reuse of existent *base* learning materials. These *base* materials can come from diverse sources of information: another kind of repositories, knowledge bases, dictionaries, etc. In this research work, we integrate techniques from Linked Open Data initiative (Semantic Web) and Text Mining for organize, search, and reuse of information from OERs. Our aim is to obtain a method to facilitate the reuse of OERs by searching OERs according to its textual content.

3 NAME: MARIO ALMAGRO CÁDIZ

TITLE: AUTOMATIC CATEGORIZATION OF ELECTRONIC HEALTH RECORDS

PHD SUPERVISORS: RAQUEL MARTÍNEZ UNANUE & VÍCTOR FRESNO FERNÁNDEZ

At present, the amount of resources allocated to the coding of Electronic Health Records (EHRs) is enormous. This task has become more complex with the recent application of the ICD-10 standard in the Spanish healthcare system. On the one hand, the CIE-10-ES coding (Spanish version of the ICD-10) gathers more than 140 thousand codes, which means a vast amount of possibilities. As for the health records, only a small number of them have so far been codified, making it difficult to develop data-based approaches. In addition, they are written in natural language, which implies the presence of jargon, spelling mistakes, incorrect syntactic structures, synonyms, acronyms and abbreviations.

This doctoral thesis aims to address the challenges posed by this task. To this end, several biomedical textual pre-processing techniques and different approaches to address the present classification problem are proposed. Supervised techniques will be studied to take advantage of the limited available data. On the other hand, different Information Retrieval techniques that use the knowledge of different medical resources will be analysed, considering ontologies such as the SNOMED and ICD manuals. Finally, other unsupervised approaches based on distributed semantics will be explored. For this purpose, a corpus of biomedical documents and medical records will be collected. Each of these approaches is expected to bring a different aspect to the pursued general solution. Therefore, different ways will be explored to combine these results and achieve an optimal and robust system that helps hospitals to save resources, thus contributing to the improvement and sustainability of the Spanish healthcare system. It is also hoped that all the results included in the thesis can be applied to Spanish-speaking countries or adaptable to other languages.

4 NAME: HERMENEGILDO FABREGAT MARCOS

TITLE: EXTRACTION OF RELATIONSHIPS BETWEEN CONCEPTS IN THE BIOMEDICAL DOMAIN

PHD SUPERVISOR: LOURDES ARAÚJO

ABSTRACT:

The detection of relationships between concepts is a particularly sensitive task in research areas such as biomedicine. The main objective of this thesis is to explore new techniques for the automatic processing of biomedical documents in order to identify biomedical entities and their relationships. Initially, the study will focus on the detection of both disabilities and diseases and their relationships. As far we know, the annotation of disabilities has not been addressed specifically in previous works. However it is an important problem that affects to a large part of the population. In addition, the notation of these entities entails special difficulties, such as the highly free language through which these entities are expressed. For the study of these objectives, different approaches based on machine learning techniques and natural language processing will be explored. Considering that the development and evaluation of some of these methods requires the existence of annotated domain corpus, one objective of this thesis is to generate a collection of annotated documents with information about disabilities and rare diseases that appear in them. This thesis will also explore the detection of negation due to the importance of it in the extraction of relationships.

5 NAME: JENNIFER HERNÁNDEZ BÉCARES

TITLE: THE WISDOM OF DIVERSE CROWDS: IMPROVING POPULARITY WEAKNESSES THROUGH DIVERSITY

PHD SUPERVISORS: JULIO GONZALO

The concept "Wisdom of crowds" is based on the idea that large groups of people are collectively smarter than individual experts when it comes to innovating, recommending items or even predicting situations. The wisdom of crowds is precisely the reason why Google Search can most of the times extract the one page that has the exact piece of information you were looking for. Google Search uses the so-called "user signals", behavioral patterns of users which Google uses to establish the rankings shown in the search results, which are directly linked to the wisdom of crowds and to popularity, since the important signals are the ones for which people are agreeing the most. However, we can find situations when the previous doesn't apply and crowds are not always wise.

As an improvement of popularity systems, we want to introduce a new concept: the wisdom of diverse crowds. This idea, contrary to popularity, involves studying how different users are between them and rating and ranking items according to the diversity of the crowd. That is, when the diversity is higher, the item is be more likely to be liked by different users with completely different profiles. Additionally, our hypothesis is that developing an algorithm that considers the diversity of the population will also mean that we could provide a solution to a very common problem in recommendation: the cold start. Finally, one of the main focuses of this thesis will be recommending new content. Here, we introduce the concept "serendipity" to avoid recommending popular items. To prove that our hypotheses are valid, we will test our ideas in two different contexts: songs recommendation and news recommendation.