

PROPUESTAS DE TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL

CURSO 2018/19

1. Particle Swarm Optimization Parameter Tuning

Director: Nik Swoboda

Descripción: Particle Swarm Optimization (PSO) is a meta-heuristic for solving optimization problems which was inspired by the flocking behavior of birds. Like many other collective intelligence techniques, an effective use of PSO requires a careful selection of the values for a number of parameters (sometimes called parameter tuning). The goal of this project is to further investigate the role of those parameters and to propose techniques for establishing their values. A number of academic publications exist which attempt to answer these questions and a careful perusal of this work is a necessary first step in this project. Based upon this existing work the student can then decide the direction she/he chooses to pursue. One suggestion would be to select data from a real-world optimization problem or collection of problems and use this as the basis of a detailed study.

2. Multi-Agent Programming Contest

Director: Nik Swoboda

Descripción: Researchers from Clausthal University of Technology organize an anual Multi-Agents Programming Contest (<http://www.multiagentcontest.org>). The 2018 contest is based upon an Agents-in-the-City scenario which consists of playing a "game" involving moving agents through a city while earning "points". Agents earn points by building and maintaining as many Wells as possible. To build wells, agents must earn money by completing Jobs or by trading items. This project involves implementing a new solution to the challenge using Collective Intelligence (CI) techniques.

3. Enabling semantic interoperability in blockchain

Director: Óscar Corcho García y Raúl García Castro

Alumno asignado: Juan Cano de Benito

4. Desarrollo de componentes de generación de grafos de conocimiento virtuales

Director: Óscar Corcho García y Raúl García Castro

Descripción: El objetivo del TFM es desarrollar componentes que den soporte a la generación de grafos de conocimiento virtuales y a su acceso, ya sea a nivel de recurso o de consulta. Los componentes enlazarán datos a partir de distintas fuentes de datos y expondrán dichos datos en los formatos recomendados por el W3C: RDF y SPARQL.

Recomendaciones: Conocimiento de desarrollo de servicios Web (REST, JSON) y Conocimiento de tecnologías semánticas (RDF, OWL, SPARQL)

4. Ontology diagram diff tool

Director: Óscar Corcho García

Descripción: The objective of this thesis work will be to track the evaluation and the diff in ontologies using snapshots of their visualisations (e.g. dot format) which are generated by the ar2dtool (and integrated in the Ontology suite). We will be able to review the changes made in a pull request (showing the two diagrams and highlight the differences).

5. Distributed SPARQL querying over heterogeneous tabular data sources

Director: Óscar Corcho García

Descripción: Design and implement a system that can query different spreadsheets (e.g. CSV files, TSV files, excels, ... etc), using SPARQL and RDB2RDF mappings.

6. Implementing a smart Query Prefetching System for SPARQL endpoints

Director: Óscar Corcho García

Descripción: Recent works point out machine learning based techniques to create more effective SPARQL endpoints. These techniques are based on predicting the structure of the incoming queries as well as catching augmented triple patterns. The experiments show that these techniques can provide an effective mechanism to avoid the execution of the incoming SPARQL queries and a fast response exploiting the cached information. This "TFM" would implement these techniques in order to achieve a prefetching layer for current SPARQL endpoint engines like Virtuoso or Fuseki.

7. Analysis on new features for a smart Query Prefetching System for SPARQL endpoints

Director: Óscar Corcho García

Descripción: Recent works point out machine learning based techniques to create more effective SPARQL endpoints. These techniques are based on predicting the structure of the incoming queries as well as catching augmented triple patterns. The experiments show that these techniques can provide an effective mechanism to avoid the execution of the incoming SPARQL queries and a fast response exploiting the cached information. This "TFM" would study new features (like time between queries) that could enhance the current results.

8. Detection of types in DBpedia based on machine learning techniques

Director: Óscar Corcho García

Descripción: Recent works have pointed out the benefits of machine learning approaches to predict the type of DBpedia resources. This "TFM" will focus on analyzing new approaches to improve the current state of the art. The supervisor will provide some suggestions but the student will have the chance to propose new approaches.

9. Generation of semantic information from natural language sources

Director: Óscar Corcho García

Descripción: With the recent advances (e.g. SyntaxNet) in the analysis of texts in natural language, the conversion of texts into RDF triples is becoming a real possibility. This "TFM" will analyze the state of the art and its application to a real use case: DBpedia. From the textual information about a given resource, many new triples will be generated. This new information must be analyzed to validate its correctness before being added to the DBpedia. This work can apply to the Google Summer of Code 2019 to get prestige and additional funding.

10. Modelo dinámico de predicción y ajuste de los precios de las habitaciones de los hoteles

Director: Antonio Jiménez Martín y Alfonso Mateos Caballero

Alumno: Andrés Herranz González

11. Simulación de sucesos raros con RESTART

Director: Antonio Jiménez Martín y José Villén Altamirano

Descripción: Este TFM supone la continuación de otro anterior en el que se implementó una herramienta software para el estudio de sucesos raros en modelos de redes (de colas o sistemas de espera) de Jackson abiertas en base al método RESTART. Se pretende finalizar el desarrollo de la herramienta con el estudio de nuevos eventos raros y su simulación.

Requisitos: Haber cursado la asignatura Métodos de simulación, C++ y conocer el framework de QT

12. Aseguramiento de calidad software en robótica aérea

Director: Martín Molina González

Descripción: El objetivo del trabajo es revisar y aplicar técnicas de aseguramiento de calidad del software que puedan utilizarse en la construcción de arquitecturas de control de vehículos robóticos aéreos que operan de forma autónoma. Se trata de arquitecturas robóticas de inteligencia artificial multinivel (por ejemplo, deliberativo, ejecutivo y reactivo). En el trabajo, el alumno realizará análisis de las diferentes propuestas existentes sobre aseguramiento de calidad y podrá aplicarlo y validarlo con ayuda del entorno de software Aerostack. Para el desarrollo del trabajo, el alumno contará con los medios y la colaboración del grupo de investigación CVAR (Visión por Computador y Robótica Aérea) de la UPM en la ETS de Ingenieros Informáticos. Este trabajo puede realizarse en un contexto de

colaboración remunerada en el propio grupo de investigación (con dedicación completa o parcial y horario flexible).

Requisitos: El alumno deberá estar familiarizado con herramientas de desarrollo informático sobre Linux.

13. Sistema de gestión de creencias para vehículos robóticos aéreos

Director: Martín Molina González

Descripción: El objetivo del trabajo fin de máster es desarrollar un sistema de gestión de creencias para vehículos aéreos robóticos aéreos (drones que operan de forma autónoma). El interés del trabajo está en encontrar una forma de conectar la información de sensores, cámaras, etc. (sistema de percepción) con los elementos simbólicos en una memoria de creencias, necesarios para planificación automática (nivel deliberativo) para que opere de forma eficiente en un vehículo aéreo. La gestión de la memoria de creencias requiere realizar tareas tales como mantener la coherencia, gestionar persistencia temporal (a corto y largo plazo), etc. En el trabajo, el alumno realizará una revisión bibliográfica para analizar las diferentes propuestas existentes sobre sistemas de este tipo en el campo de la robótica y realizará el diseño y programación experimental en C++ (sobre Linux) de un prototipo para vehículos aéreos con ayuda del entorno de software Aerostack para ser evaluado con drones en vuelos reales. Para el desarrollo del trabajo, el alumno contará con los medios y la colaboración del grupo de investigación CVAR (Visión por Computador y Robótica Aérea) de la UPM en la ETS de Ingenieros Informáticos.

Requisitos: El alumno deberá estar familiarizado con herramientas de desarrollo informático sobre Linux.

14. A multi-criteria approach for assigning the power index in voting systems

Director: Jacinto González Pachón

Descripción: In many voting systems, decision makers (DMs) are not treated equally. This is the case of participatory decision-making problems involving stakeholders, where, if one of them has more shares of stock, then her/his vote is weighted more heavily (e.g., the case of the European Union institutions, in which larger member countries obtain more votes). In this proposal, the interest is on how to assign a system of weights in a weighted voting system, in order to quantify group versus individual trade-offs in voting situations

15. Combining ethical principles in participatory decision making: In search for a compromise

Director: Jacinto González Pachón

Descripción: When we are dealing with participatory decision-making problems, hypothetical conflicts between some ethical principles should not be ignored or covered up, but encouraged. The question is “is it possible to quantify the degree of conflict between the achievements of these principles in a particular participatory decision making problem? Alternatively, “is it possible to compute the sacrifice degree of some principles in order to reach a final solution? In this proposal, the interest is on an analytical framework for dealing with conflictive basic ethical principles under the satisficing paradigm.

16. Valoración de películas basada en Deep Learning y Computación Afectiva

Director: Jacinto González Pachón y Emilio Serrano

Descripción: El Deep Learning es un subcampo del aprendizaje automático basado en el uso de redes neuronales artificiales que, mediante una jerarquía de capas con unidades de procesamiento no lineal, aprenden abstracciones de alto nivel para datos. El Deep Learning ha supuesto un avance histórico en algunos de los problemas más complejos de la IA como la visión por computador (p. ej. en la conducción autónoma de los coches Tesla) y el reconocimiento automático del habla (p. ej. en Google Assistant y Siri). En esta tesis de máster se explorarán métodos de Deep Learning que combinen: la extracción de información semántica de vídeos; la computación afectiva para detectar las emociones en películas; y el estudio de correlaciones entre esta información y la puntuación dada por los espectadores a las películas en la Internet Movie Database (IMDb).

17. Una revisión sistemática del uso de tecnologías basadas en agentes para el desarrollo del Internet de las Cosas

Director: Javier Bajo Pérez y Emilio Serrano

Descripción: El Internet de las Cosa (IoT) plantea un mundo en el que todos los objetos físicos están etiquetados, identificados y con la capacidad de inter-operar en la infraestructura de Internet. Pero más allá de esta visión inicial, el IoT es un campo multidisciplinar que está permitiendo a las máquinas percibir el mundo e interactuar de manera transparente con este. Algunas de sus aplicaciones incluyen monitorización de salud, el control industrial de procesos como la agricultura, los entornos y ciudades inteligentes, etcétera. Las tecnologías basadas en agentes, como los sistemas multiagente y la simulación multiagente, se adaptan naturalmente al diseño de un sistema IoT, especialmente cuando se busca dotar al sistema de un comportamiento inteligente. En esta tesis de máster se plantea la realización de una revisión sistemática del uso de tecnologías basadas en agentes para el desarrollo del IoT; caracterizándose una revisión sistemática por el uso de métodos explícitos y rigurosos para identificar, evaluar, y sintetizar estudios relevantes en una temática particular.

18. Aprendizaje por Refuerzo mediante Deep Learning para las Ciudades Inteligentes

Director: Javier Bajo Pérez y Emilio Serrano

Descripción: El aprendizaje por refuerzo permite que agentes aprendan automáticamente a tomar una serie de acciones sobre un entorno de manera que se maximiza una recompensa. El Deep Learning, un subcampo del aprendizaje automático basado en el uso de redes neuronales artificiales, ha sido ampliamente utilizado en los últimos años como tecnología para resolver problemas de aprendizaje por refuerzo. Por otro lado, las ciudades inteligentes son áreas urbanas que usan diversos datos recolectados de los ciudadanos, dispositivos, y otros activos; que son procesados y analizados para monitorizar y gestionar los recursos de la ciudad de manera inteligente y eficiente. Esta tesis de máster plantea: el estudio de aplicaciones del aprendizaje por refuerzo en los espacios y ciudades inteligentes que se hayan realizado en la literatura especializada, la propuesta de nuevas aplicaciones

no exploradas previamente, y la experimentación con esta tecnología en simulaciones para apoyar la hipótesis de que su uso es útil y viable.

19. Aprendizaje basado en experiencias pasadas para el soporte de las ciudades inteligentes

Director: Javier Bajo Pérez y Emilio Serrano

Descripción: Las ciudades inteligentes son áreas urbanas que usan diversos datos recolectados de los ciudadanos, dispositivos, y otros activos; que son procesados y analizados para monitorizar y gestionar los recursos de la ciudad de manera inteligente y eficiente. La cantidad de servicios que pueden ser asistidos por la Inteligencia Artificial en estas ciudades es innumerable: transporte, plantas nucleares, suministro de agua, la gestión de residuos, hospitales, bibliotecas, escuelas, etcétera. El aprendizaje y el razonamiento automático son claves en el desarrollo de las ciudades inteligentes. El razonamiento basado en casos o case-based reasoning (CBR) es una tecnología que, esencialmente, permite resolver problemas basándose en la solución que se dio previamente a problemas similares. Una de las grandes ventajas del CBR sobre otros métodos es su gran cercanía al modo de razonar y aprender de los seres humanos, así como su capacidad de dar explicaciones comprensibles sobre sus decisiones. Esta tesis de máster plantea: el estudio de aplicaciones del CBR en espacios y ciudades inteligentes que se hayan realizado en la literatura especializada, la propuesta de nuevas aplicaciones no exploradas previamente, y el diseño de un sistema de aprendizaje basado en experiencias pasadas para el soporte de estos espacios.

20. Automated tagging and extraction of information from breast cancer radiology reports

Director: Miguel García Remesal

Descripción: This project is targeted at reviewing, applying (and ideally creating) novel techniques to identify and extract relevant information from breast cancer radiology reports. This includes identifying mentions to relevant entities and their associated features in the radiology reports, to create structured clinical records that can be used to facilitate physicians their daily work. We will work with real-world reports, collaborating with radiologists from one of the most prestigious hospitals in Madrid.

21. Temporal evolution of European Commission-funded research projects in health informatics

Director: Miguel García Remesal

Descripción: The proposed work is aimed at creating an informatics tool that will facilitate the automated creation of graphs that characterise the main research topics that were funded in a given EC health informatics call and their temporal evolution. The student will resort to a wide range of text mining techniques to automatically analyze and extract useful information from CORDIS records to build the graphs.

21. Methods for disambiguating author names in scientific publications

Director: Miguel García Remesal

Descripción: The student will review the available literature to find methods and tools targeted at disambiguating author names in scientific publications, recorded patents, etc. These methods will be applied to a set of selected scientific papers and recorded patents to evaluate and compare their performance. An optional (desirable) objective of this work would involve the creation of a novel method for author name disambiguation that outperforms other existing methods.

22. Búsqueda de comunidades de relaciones homogéneas en grafos ponderados. Aplicación a la detección de tramas de blanqueo de capitales

Director: Alfonso Mateos Caballero

Descripción: Un ejemplo de trama de blanqueo de capitales está caracterizada por la construcción de un entramado empresarial cuyos objetivos consisten en mover dinero aparentemente limpio entre las empresas de la trama para blanquear en cada transacción una parte del dinero procedente de los negocios ilícitos.

Por ejemplo, si la empresa A tiene como único cliente a la empresa B, y la empresa B tiene como único cliente a la empresa A, cabría preguntarse dónde está el beneficio de ambos. Si en lugar de dos empresas el procedimiento se repite entre n empresas que no tienen clientes fuera del grupo, el planteamiento es exactamente el mismo, y si dichas empresas en lugar de estar completamente aisladas del exterior, tienen algunas relaciones comerciales con otras empresas, pero la mayor parte de sus transacciones se realizan en el interior de la trama, la situación es muy parecida.

Por tanto, un tipo de red posible de blanqueo de capitales estaría constituida por un conjunto de empresas con una enorme cantidad de relaciones comerciales entre ellas en relación a las relaciones que se tienen con el exterior.

Puesto que las relaciones comerciales deben declararse a la Hacienda Pública, y puesto que de lo contrario, el dinero sucio no sería transformado, resulta que podemos construir un grafo de relaciones comerciales entre empresas y por tanto, en dicho grafo podemos buscar comunidades de empresas que obedezcan a esta topología: una gran cantidad de relaciones comerciales entre las empresas de la comunidad y un relativo aislamiento con respecto al exterior.

Sin embargo, resulta además de interés plantear este problema en términos cuantitativos en lugar de en términos topológicos. No nos importaría tanto la cantidad de relaciones dentro/fuera sino la cantidad de dinero que se mueve en el interior de las tramas con respecto a la cantidad de dinero que entra o sale de la trama.

Visto así, el problema puede plantearse a partir de una función de convolución sobre los nodos de un grafo inducido a partir del grafo original ponderado por la cuantía de la transacción comercial. La solución sería similar a la detección de bordes en una imagen que separan las diferentes regiones formadas por los colores de los objetos de dicha imagen. Nuestro problema radica en optimizar la construcción de ese grafo inducido cuyos detalles se mostrarán más detenidamente al estudiante del TFM.

La solución se plantea a un problema de interés en la Agencia Estatal de la Administración Tributaria.

23. Versión paralela del algoritmo de clustering DBSGRAPH. Aplicaciones en el sector turístico

Director: Alfonso Mateos Caballero

Descripción: El algoritmo DBSGRAPH es un algoritmo de clustering que mejora a la mayoría de los algoritmos clásicos (k-means, fuzzy c-means o DBSCAN) asumiendo sus virtudes y evitando sus defectos. Este algoritmo está basado en la construcción de un grafo para distintos niveles de un parámetro, la detección de componentes conexas de dicho grafo, la optimización del nivel del parámetro y el uso de una función de centralidad para establecer una función de pertenencia al cluster. El algoritmo además es capaz de separar regiones que no son linealmente separables.

El problema consiste en definir y desarrollar una versión paralelizada del algoritmo que acelere su convergencia.

Este algoritmo se plantea para resolver un problema de interés en el Ministerio de Comercio, Industria y Turismo.

24. Desarrollo de una metaheurística para completar el algoritmo k-Graph de detección de tramas cohesionadas en redes sociales. Aplicaciones a la detección de tramas de blanqueo de capitales

Director: Alfonso Mateos Caballero

Descripción: El algoritmo k-graph es un algoritmo de detección de comunidades altamente cohesionadas en redes complejas, y ha dado excelentes resultados en la detección de tramas de blanqueo de capitales. Este algoritmo es la versión topológica de la solución planteada en el Problema 1 y por tanto es de interés para la Agencia Estatal de la Administración Tributaria.

El problema del algoritmo es que la condición de comunidades cohesionadas es demasiado restrictiva, y se requiere implementar una metaheurística que suavice esta restricción pero que siga manteniendo un cierto nivel de cohesión y aislamiento para las tramas resultantes.

25. Detección de fraude sobre grandes datos fiscales mediante técnicas analíticas. Una revisión de usos de herramientas de minería de datos aplicadas sobre un caso real

Director: Alfonso Mateos Caballero

Descripción: La recaudación de impuestos se ha asociado a un gran número de fraudes, lo que supone un reto a la hora de detectarlos. Esto se debe al crecimiento en el tamaño de los datos y también a la ausencia de procesos de negocio totalmente automatizados. La Agencia Tributaria Española (AEAT) alberga enormes cantidades de datos que necesitan mecanismos complejos para extraer información fiscal útil. Hasta la actualidad, las herramientas de minería de datos y los sistemas de inteligencia artificial han generado grandes avances tecnológicos en la forma en que se detecta el fraude en el sector de la Administración Tributaria. La capacidad y rendimiento de tales métodos en la presentación

de informes y la extracción de conocimientos y patrones de las bases de datos muestran que el conocimiento en el sector tributario es muy importante. En este punto, nuestro objetivo es realizar una revisión bibliográfica de las diferentes herramientas y métodos de minería de datos e inteligencia artificial aplicados hasta el momento para la detección de fraude fiscal. Una aplicación de un caso real de estudio con datos fiscales nos ayudará a entender las ventajas y desventajas de cada uno de ellos a la hora de detectar el fraude. Finalmente trataremos de implementar alguna de las herramientas sistemáticas utilizadas en la detección de fraude fiscal aplicada a datos abiertos.

26. Técnicas MCDM para la restauración de balsas de fosfoyesos

Director: Antonio Jiménez Martín y Danyl Pérez Sánchez (CIEMAT)

Descripción: El objetivo del TFM es aplicar la herramienta de MCDA para la toma de decisiones, en el escenario de residuos NORM correspondiente a las balsas de fosfoyesos de Huelva, teniendo en cuenta aspectos económicos, sociales, medioambientales y radiológicos. Existen dos opciones de restauración del emplazamiento, la primera es la remediación in situ, utilizando el acondicionamiento de las balsas y cobertura vegetal, otra alternativa es la remoción de estos residuos a un vertedero de residuos convencional y su reutilización en otras aplicaciones como material para emendar suelos, construcción, carreteras, etc.

27. Inducción de los esquemas narrativos eventos

Director: Igor Boguslavskiy

Descripción: Uno de los enfoques para modelar la comprensión de Lenguaje Natural es el enfoque “top-down” que utiliza representaciones de alto nivel como scripts. Los scripts son las representaciones estructuradas de los eventos, sus relaciones causales y sus participantes. Los scripts son útiles para las aplicaciones como resumen automático, resolución de correferencia y sistemas pregunta-respuesta. El problema serio de los scripts es que son construidos manualmente. Las técnicas de aprendizaje automático permiten inducir las representaciones similares - las secuencias estereotípicas de los eventos (esquemas narrativos de eventos - ENE) automáticamente. Un ejemplo de ENE es: “cometer un crimen” – “arrestar” – “juzgar” – “condenar”. La tarea de TFM consiste en aprender los ENEs de los textos libres.

Condiciones: Haber cursado las asignaturas de Ingeniería Lingüística y Aprendizaje Automático o poseer los conocimientos que se imparten en las mencionadas asignaturas.

28. Generación de artículos en español

Director: Igor Boguslavskiy

Descripción: El objetivo del trabajo es la generación automática de artículos definidos e indefinidos durante la generación del texto español. El guion típico en que surge esta tarea es la traducción automática de una lengua fuente que no disponga de artículos o tenga las reglas de su uso distintas del español. Un ejemplo del uso distinto de los artículos entre el español y el inglés es: *Soy estudiante en la Universidad de Oxford – I am a student at Oxford University*. Habrá que utilizar un corpus bilingüe alineado para entrenar un clasificador.

Condiciones: Para realizar esta tesis es recomendable haber cursado las asignaturas de Aprendizaje Automático y Ingeniería Lingüística.

29. Aplicación de Long-Short Term Memories

Director: Daniel Manrique Gamo

Alumno: **Juan Antonio Carrillo Segura**

30. Factorización de Darboux de matrices por bandas

Director: Daniel Manrique Gamo

Alumno: **David Griñán Martínez**

31. Sistema evolutivo híbrido para la construcción de redes de neuronas

Director: Daniel Manrique Gamo

Alumno: **Carlos Vázquez Losada**

32. Seguimiento ocular para análisis del comportamiento

Director: Javier de Lope Asiaín

Descripción: El análisis del movimiento de los ojos puede ayudar a determinar qué actividades puede estar realizando una persona. Además de la información obtenida por el movimiento de los ojos, también pueden utilizarse otros sensores para que registren con más calidad o detalle la actividad que se esté realizando. Por ejemplo, otras cámaras que se encarguen de otras partes del cuerpo o sensores de actividad bioeléctrica cerebral.

33. Análisis de datos de sistemas de captura de movimiento

Director: Javier de Lope Asiaín

Descripción: El objetivo es la creación de modelos de movimiento del cuerpo humano a partir de datos tomados con sistemas inerciales de captura de movimientos y posible comparación de resultados con datos tomados con sistemas clásicos basados en imágenes.

34. Aprendizaje de Redes Bayesianas con Datos Nominales, Ordinales y Continuos

Director: Juan Antonio Fernández del Pozo

Descripción: Se trata de construir-aprender modelos de red Bayesiana para datos en escalas diferentes. Tras la preparación de los datos y aprendizaje se hará una evaluación de la calidad de los modelos.

Requisitos: Conocimientos de R: entorno, paquetes y programación
Asignaturas recomendadas: Redes Bayesianas, Aprendizaje Automático, Sistemas de Ayuda a la Decisión

35. Análisis de Redes Bayesianas Mediante Clustering de Distribuciones de Probabilidad Condicionada

Director: Juan Antonio Fernández del Pozo

Descripción: Se trata de categorizar las tablas de probabilidad condicionada para facilitar la asignación de probabilidades y la validación de modelos complejos que incluyen tablas muy grandes.

Requisitos: Conocimientos de R: entorno, paquetes y programación
Asignaturas recomendadas: Redes Bayesianas, Aprendizaje Automático, Sistemas de Ayuda a la Decisión

36. Transformación de Clasificadores Bayesianos Discretos en Binarios

Director: Juan Antonio Fernández del Pozo

Descripción: Se trata de realizar un algoritmo para transformar un modelo de red Bayesiana con variables en dominio discreto a dominio binario y estudiar la evaluación e interpretación de la inferencia con variable 0/1.

Requisitos: Conocimientos de R: entorno, paquetes y programación
Asignaturas recomendadas: Redes Bayesianas, Aprendizaje Automático, Sistemas de Ayuda a la Decisión

37. Reconocimiento de landmarks basado en el método bolsa de palabras visuales (bag of words) para la navegación autónoma mediante mapas Topológicos visuales

Director: Darío Maravall Gómez-Allende

Descripción: El objetivo es aplicar el método “bolsa de palabras visuales (bag of words) para el reconocimiento robusto de landmarks en mapas topológicos visuales para la navegación autónoma en interiores.

Requisitos: Haber cursado con buena calificación la asignatura Robots Autónomos del curso 2017/ 2018, ya que es necesario tener conocimientos suficientes de los fundamentos de la navegación autónoma mediante Mapas Topológicos visuales

38. Reconocimiento de landmarks visuales basado en Deep learning+CNN para la navegación autónoma mediante Mapas Topológicos visuales

Director: Darío Maravall Gómez-Allende

Descripción: El objetivo es aplicar el método “bolsa de palabras visuales (bag of words) para el reconocimiento robusto de landmarks en mapas topológicos visuales para la navegación autónoma en interiores.

Requisitos: Haber cursado con buena calificación la asignatura Robots Autónomos del curso 2017/ 2018, ya que es necesario tener conocimientos suficientes de los fundamentos de la navegación autónoma mediante Mapas Topológicos visuales.

39. Gene regulatory Bayesian networks in the brain

Director: Concha Bielza Lozoya y Pedro Larrañaga Múgica

Alumno: Nicolás Bernaola Álvarez

40. Clasificadores bayesianos en la predicción del Alzheimer

Director: Concha Bielza Lozoya y Pedro Larrañaga Múgica

Alumno: Javier Ramos Fernández

41. Aprendizaje automático en documentos de abogacía

Director: Concha Bielza Lozoya y Pedro Larrañaga Múgica

Alumno: Franklin Parrales Bravo

42. Metodología para puntuación automática de textos transcritos desde voz en lengua española

Director: Jesús Cardeñosa Lera

Alumno: Laura Gálvez Jiménez

Descripción: Uno de los problemas de la capacidad de saber de que se está hablando y posteriormente saber que se dice de algo en concreto en una noticia en TV o radio es que los sistemas de transcripción recogen tan solo la representación escrita de palabras sin ningún signo de puntuación y en ocasiones de tildes. Eso permite solamente la búsqueda basada en la identificación de cadenas de caracteres. Cualquier trabajo que quiera ir más allá, como saber de que trata una noticia requiere una información adicional que los sistemas actuales no proveen en absoluto para la lengua española. Esta tesis de máster pretende llegar a establecer una metodología de tratamiento de textos transcritos a partir de voz para por medio de puntuación pueda saberse algo más como la morfología de las palabras o pasos más allá llegando a identificar frases subyacentes que permitan identificación de datos o establecimiento de relaciones entre conceptos presentes.

43. Modelo para la identificación automática de argumentos de opiniones en foros o redes sociales

Director: Jesús Cardeñosa Lera

Alumno: Daniel Blázquez Cabezas

Descripción: Las redes sociales proporcionan información rápida sobre cuestiones como análisis de sentimiento. Sin embargo el saber la polaridad de una opinión no tiene una gran validez si no se acompaña de las razones que han llevado a la misma. La extracción de argumentos basándose en criterios lingüísticos es casi la única opción para establecer razones de una opinión (argumentos) y con ello establecer pautas de cambio sobre productos, marcas o servicios prestados por diferentes tipos de empresas o instituciones. Es quizás una de las informaciones que más cuesta obtener a los responsables de distintos tipos de servicios. Se pretende en este trabajo definir un modelo que será ratificado experimentalmente para identificar y extraer los argumentos de opiniones

44. Temporal Expressions and Events in the legal domain

Director: Víctor Rodríguez Doncel y Óscar Corcho García

Work description: Legal documents (EU directives, national laws, judgements) contain references to temporal expressions (dates, times, periodic events, intervals) and events (a hearing, a robbery, a lodgement of a complaint...). These textual references are not made explicit, hampering the navigation of the documents, preventing the development of advanced search engines and decreasing the performance of information analysis algorithms. Being able to find them and deciding among the most relevant ones would help also to make legal documents easier to represent and understand, generating for instance summaries or timelines.

The person taking this assignment would help in the endeavour of annotating and evaluating the presence of these temporal entities in legal documents. The person will help creating new resources to facilitate this for English and/or Spanish, and other languages if possible, and can also collaborate in the algorithms for tagging and visualization.

Desired skills: The person taking this assignment would ideally have interest in NLP technologies and a fair knowledge of either Java or Python. Specific skills on NLP toolkits are not required but would be a plus.

45. Analysis of mobility patterns of elderly citizens

Director: Víctor Rodríguez Doncel y Óscar Corcho García

Work description: Elderly citizens exhibit a decay in their activity as they grow older. This decay can be observed in their mobility routines, and public transportation authorities can observe it. The objective of this TFM consists of modeling mobility behaviour patterns in elderly citizens and devise the appropriate algorithms to detect anomalies –such as elderly citizens whose use of mass transit decay faster than expected. Datasets to be analyzed include transportation data from public authorities as well as different open data sources.

Desired skills: Algorithm design (in R/Matlab or equivalent), programming (Python/Java), basic user interfaces (web application).

46. Análisis de imágenes en dispositivos móviles

Director: Luis Baumela Molina

Descripción: Se desarrollarán algoritmos para análisis de imágenes en tiempo real sobre dispositivos móviles con el fin de detectar, describir y emparejar líneas.

Requisito imprescindible haber cursado satisfactoriamente la asignatura de Visión por Computador.

47. Análisis de imágenes para interacción hombre-máquina

Director: Luis Baumela Molina

Descripción: Se desarrollarán sistemas de visión por computador que analizan imágenes de cámaras que miran a personas y obtienen información sobre el movimiento y la configuración del cuerpo, atributos tales como identidad, expresiones del rostro, etc. Utilizaremos fundamentalmente aproximaciones basadas en Deep Learning.

Requisito imprescindible haber cursado satisfactoriamente la asignatura de Visión por Computador.

48. Annotating Contracts with Linked Terminologies

Director: Óscar Corcho García

Descripción: Contracts are agreements which as any other document in the legal domain, contain words and expressions (terms) that have a specific meaning in the context they are used, often with legal implications.

Big companies have to deal with different types of contracts, and they would benefit from AI tools that would help them in processing those contracts (annotating and classifying them, extracting important information from them, etc.).

In this TFM, the student will have to work in the annotation of contracts with terminologies (collections of terms used contracts) in RDF. The person taking this assignment would harvest and/or create legal terminologies in the domain of contract law, convert them into RDF (according to existing models) and link them to other terminologies in the Linked Data cloud. Then, he or she would work in the annotation of real contracts with those terminologies.

Desired skills: The person taking this assignment would ideally have interest in NLP technologies and a fair knowledge of RDF. Specific skills on NLP toolkits are not required but would be a plus.

49. Computación evolutiva para la estimación de parámetros en el simulador GRO

Director: Alfonso Rodríguez Patón

Descripción: ¿Qué es GRO? Es un lenguaje para la programación, modelado, especificación y simulación de biocircuitos en sistemas multicelulares que se emplea en Biología Sintética. GRO simula y modela el crecimiento de bacterias en microcolonias. Está especialmente diseñado para el prototipado de comportamientos multicelulares, el estudio de la propagación de información intercelular y el procesamiento de esta información de manera distribuida. GRO es un sistema basado en agentes (cada célula es un agente o individuo) con un potente interfaz gráfico que permite analizar y observar la emergencia de complejos procesos multicelulares a partir de la especificación de simples reglas de interacción local entre células o biomoléculas. El código es abierto y el LIA lleva años desarrollando nuevas funcionalidades en este entorno de programación celular.

Web del simulador gro:

<http://depts.washington.edu/soslab/gro/>

Referencias sobre GRO:

M. Gutiérrez, P. Gregorio-Godoy, G. Pérez del Pulgar, L.E. Muñoz, S. Sáez, and A. Rodríguez-Patón, "A New Improved and Extended Version of the Multicell Bacterial Simulator gro" ACS Synthetic Biology, April, 2017.

S.S. Jang, K.T. Oishi, R.G. Egbert, and E. Klavins, "Specification and simulation of multicelled behaviors", ACS Synthetic Biology, July, 2012.

¿En qué consiste este TFM?

El trabajo se centrará en la mejora de alguna funcionalidad de GRO, en concreto, permitir que gro se pueda ejecutar por lotes en modo "batch" para poder lanzar diferentes experimentos in silico de manera automática siguiendo algún algoritmo de optimización (alguna técnica de computación evolutiva como Simulated Annealing, Particle Swarm Optimization o Ant Colony Optimization) que ajuste los parámetros del modelo simulado a datos reales. También se mejorará la gestión y almacenamiento de los datos generados en las simulaciones.

¿Hay que saber mucha biología para este TFM? No. Prácticamente nada. Es un proyecto de programación de manera que es más que suficiente con lo aprendido durante la asignatura Biología Programable. Se trabajará en el grupo LIA en colaboración con biotecnólogos, físicos e informáticos con experiencia previa en el simulador GRO.

Tareas en este TFM:

1. Aprender el funcionamiento del simulador gro (escrito en C++).
2. Programar la infraestructura de entrada/salida de gro para permitir la ejecución por lotes.
3. Programar algún método de computación evolutiva aplicado a optimización (Particle Swarm Optimization, Simulated Annealing o Ant Colony Optimization) para el ajuste de las simulaciones a datos reales. Se emplearán los algoritmos desarrollados en nuestro grupo en el paquete de R denominado EvoPER.

Prestes García A, Rodríguez-Patón A. 2016. EvoPER -- An R package for applying evolutionary computation methods in the parameter estimation of individual-based models implemented in Repast.

PeerJ Preprints 4:e2279v1 <https://doi.org/10.7287/peerj.preprints.2279v1>

Conocimientos previos recomendados: C++ (medio/alto nivel). Conocimiento de técnicas de computación evolutiva.

50. Sin título

Director: Alfonso Rodríguez Patón

Alumno: Javier París

51. Buscador de patrones en señales temporales complejas con modelos ocultos de Markov

Director: Óscar Corcho García

Empresa: Mirubee

Remuneración: 600 €/mes (esfuerzo estimado de 5 horas al día durante unos 3 meses)

Lugar de realización del TFM: ETSI Informáticos, visitando una vez por semana las oficinas de Mirubee en la calle del Aviador Zorita.

Descripción: Existen hoy día sistemas para discriminar consumos eléctricos sencillos dentro de una señal compleja mediante algoritmia, sin necesidad de usar sub-medición física, usando sólo un medidor general en un establecimiento. Los sistemas actuales detectan principalmente consumos de tipo on-off, singulares y de poca duración (planchas, calefactores, etc.). El reto tecnológico que se plantea es el de poder discriminar señales más complejas utilizando modelos ocultos de Markov (HMM), aplicándolo a equipos consumidores multinivel (varios niveles de potencia) programables, así como discriminando el consumo simultáneo de múltiples cargas menores y de duración incierta como el caso del alumbrado.

Keywords: Procesamiento de señales temporales y desarrollo de algoritmos

52. Deep Learning out of the (black) box

Director: Luis Baumela Molina

Alumno: Jorge Marco Blanco

53. Elaboración de preguntas de competencia basada en patrones

Directora: María del Carmen Suárez de Figueroa

Descripción: Actualmente, las preguntas de competencia (requisitos ontológicos) se elaboran de manera manual sin ninguna pauta específica. Esto supone una dificultad añadida en el establecimiento de los requisitos que debe cumplir una ontología. Por este motivo, se propone el desarrollo de un software que sirva de apoyo en la creación de preguntas de competencia y que base dicha creación se base en una colección de patrones lingüísticos.

Condiciones para el candidato: Conocimientos de ingeniería ontológica

54. Identificación y transformación de oraciones complejas

Directora: María del Carmen Suárez de Figueroa
Alumna: Clara Cerrato

Descripción: Las oraciones compuestas son difíciles de comprender por parte de ciertos colectivos (personas con discapacidad cognitiva, extranjeros con conocimientos limitados del idioma, entre otros). Por este motivo, la identificación de este tipo de oraciones y su transformación en oraciones simples puede resultar de gran ayuda. Se plantea en esta TFM una aplicación que sea capaz de identificar diferentes tipos de oraciones complejas en castellano y que proponga una serie de transformaciones para su simplificación.

55. Visualización de redes bayesianas para big data

Director: Concha Bielza Lozoya y Pedro Larrañaga Múgica
Alumno: Sergio Paniego Blanco

56. Minería de textos en contratos corporativos

Director: Concha Bielza Lozoya y Pedro Larrañaga Múgica
Descripción:

57. Resolución táctica del problema de asignación de controladores aéreos a puestos de control

Director: Antonio Jiménez Martín y Alfonso Mateos Caballero
Alumno: Sergio Flórez Vallina

58. Aplicación de técnicas de Inteligencia Artificial al entorno ATM (Air Traffic Management)

Director: Antonio Jiménez Martín y Álvaro Rodríguez Sanz (ETSI Aeronáuticos)
Alumno: Rohit Uttamchandani Dadlani
Alumno: Jonathan Eduardo Tito Ontaneda

59. Aplicación de técnicas de Inteligencia Artificial al cálculo de la carga de trabajo de controladores aéreos

Director: Alfonso Mateos Caballero y Álvaro Rodríguez Sanz (ETSI Aeronáuticos)
Alumno: José Joaquín Peralta Abadía

60. Early detection of depressive states from text analysis

Director: Óscar Corcho García

Alumno: **Carla Fernández González**

61. Using recommendation systems to enhance local citizen proposals

Director: Óscar Corcho García

Alumno: **Rubén López Lozoya**
