A9. Datos temporales y complejos

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Másteres Universitarios

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.

DATOS GENERALES

Breve descripción

En esta asignatura se aprenderá a analizar series temporales con el objeto de predecir su comportamiento futuro. Para ello, se trabajará en aplicaciones reales con series temporales de consumo de energía.

Además, se estudiarán los flujos de datos, conocidos como data streams, cuya principal característica es que su análisis debe realizarse en tiempo real. También se estudiará la clasificación multi-etiqueta, en la que los datos se clasifican con varias etiquetas diferentes a la vez.

Por último, se estudiará la cuantificación para que sepas estimar la distribución de las clases que contiene una muestra.

Título asignatura

A9. Datos temporales y complejos

Código asignatura

102472

Curso académico

2022-23

Planes donde se imparte

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

Créditos ECTS

4,5

Carácter de la asignatura

OPTATIVA

Duración

Anual

Idioma

Castellano

CONTENIDOS

Contenidos

Se pretende unir tanto los algoritmos de aprendizaje que usan como entradas tipos de datos estructurados (como los grafos y las series temporales que serían un caso particular) y aquellos algoritmos que aprender funciones cuya salida son tipos de datos estructurados.

- Datos temporales: Series temporales, Flujos de datos.
- Aprendizaje parcialmente supervisado: semi-supervisado, etiquetado positivo, multi-instancia.
- Salidas estructuradas: multi-etiqueta, Aprendizaje con grafos, Conditional Random Fields (CRF).

Unidades

Módulo 1. Series temporales
Tema 1. Introducción
Tema 2. Predicción de series temporales
Tema 3. Outliers

Módulo 2. Data Streams
Tema 1. Introducción
Tema 2. Metodología
Tema 3. Minería de datos sobre data streams

Módulo 3. Clasificación multi-etiquetas
Tema 1. Introducción
Tema 2. Algoritmos
Tema 3. Dependencias entre etiquetas

Módulo 4. Cuantificación
Tema 1. Introducción
Tema 2. Algoritmos

RESULTADOS DE APRENDIZAJE Y DE FORMACIÓN

Generales

CG1 - Entender los conceptos, los métodos y las aplicaciones de la inteligencia artificial.

CG2 - Evaluar nuevas herramientas computacionales y de gestión del conocimiento en el ámbito de la Inteligencia Artificial.

CG3 - Gestionar de manera inteligente los datos, la información y su representación.

Específicas

CE2 - Aplicar las técnicas de aprendizaje automático utilizando la metodología de validación y presentación de resultados más apropiada en cada caso.

CE5 - Analizar las fuentes documentales propias del ámbito de la investigación en Inteligencia Artificial para poder determinar cuáles de ellas son relevantes en la resolución de problemas concretos.

PLAN DE APRENDIZAJE

Actividades formativas

A1 - **Sesiones presenciales virtuales (clases en vídeo)**: visionado inicial del material audiovisual que constituye las lecciones de la asignatura. Se asume 2 veces el tiempo real de vídeo, puesto que el estudiante deberá parar, repetir, etc. algunas secuencias (10 horas).

A2 - **Trabajos individuales**: realización de tests, realización de ejercicios, realización de un trabajo de investigación y realización de un examen oral (30,5 horas).

A3 - **Trabajo autónomo**: estudio del material básico, lecturas complementarias y otros contenidos (60 horas).

A4 - **Foros y chats**: lanzamiento, lectura y contestación de cuestiones y temas para la discusión general (6 horas).

A5 - **Tutorías**: consultas y resolución de dudas, aclaraciones, etc. (6 horas).

Puede consultar en este enlace el **Cronograma de Carga de Trabajo**.

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

La evaluación de la asignatura se basará principalmente en los conocimientos teóricos adquiridos a través de los videos y transparencias, así como en los conocimientos prácticos obtenidos a través de la realización de un trabajo de investigación relacionado con uno de los temas de la asignatura.

La nota final, entre 0 y 10 puntos, se calculará teniendo en cuenta tres actividades evaluables:

E1 - **Valoración de los cuestionarios de evaluación** (30% de la calificación). Cada uno de los temas de la asignatura tendrá varias pruebas tipo test que deberán ser realizadas por los estudiantes tras el estudio de los temas. Cada test será evaluado de 0 a 10 puntos y el estudiante dispondrá de dos intentos para realizarlos. La nota final de cada test será la nota media de los dos intentos.

E2 - **Valoración de la participación en foros y chats** (10% de la calificación). Se valorará el nivel de participación/debate de los estudiantes que contará para la nota final.

E3 - **Valoración de los trabajos individuales** (60% de la calificación). Un examen oral (15% de la calificación) y un trabajo de investigación (45% de la calificación). Se realizará un examen oral, de unos 10 minutos de duración vía Skype o tecnología similar, de los bloques Multi-etiqueta y Cuantificación. En dicho examen se discutirán aspectos teóricos fundamentales de los clasificadores multi-etiqueta y los algoritmos de cuantificación. El objetivo es evaluar si el estudiante domina o no los conceptos de dichos métodos de aprendizaje. Los profesores propondrán un trabajo de investigación para cada uno de los bloques Series Temporales y Aprendizaje en Flujos de Datos. El estudiante deberá elegir un trabajo de entre los dos propuestos. El trabajo debe hacerse de forma individual y se entregará a través de la plataforma de docencia online en el plazo establecido según el calendario del máster.

Calendario de exámenes

Para la **convocatoria ordinaria**, habrá 3 fechas de entrega de trabajos final de curso. Los alumnos podrán entregar sus trabajos en cualquier momento, pero sólo en estas fechas se recogerán y evaluarán los que se hayan entregado. Las fechas serán:

• 13/01/23
• 17/03/23
• 31/05/23

Habrá una **convocatoria extraordinaria** en todas las asignaturas. Para su evaluación, la fecha límite para la entrega de trabajos será:

• 14/07/23

Las actas de la convocatoria ordinaria se cerrarán en julio de 2023 y las de la convocatoria extraordinaria en septiembre de 2023.

PROFESORADO

Profesor responsable

Troncoso Lora, Alicia

DOCTORA EN INGENIERÍA INFORMÁTICA.
CATEDRÁTICA DE UNIVERSIDAD.
UNIVERSIDAD PABLO DE OLAVIDE DE SEVILLA.

Profesorado

Riquelme Santos, José Cristóbal

Catedrático de Lenguajes y Sistemas Informáticos
Universidad de Sevilla

Melgar García, Laura

Doctora en Ingeniería Informática - especialidad Inteligencia Artificial.
Profesora Ayudante Doctora.
Universidad Politécnica de Madrid.

Del Coz Velasco, Juan José

Profesor Titular de Ciencias de la Computación e Inteligencia Artificial
Universidad de Oviedo

Fernández Díaz, Miriam

Teaching Assistant
Ingeniero en Innovación
Arcelormittal Global R&D España
Universidad de Oviedo

HORARIO

Horario

Todas las asignaturas estarán en la plataforma a disposición de los estudiantes desde octubre hasta julio.

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Series Temporales

Ian H. Witten and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann, June 2005.

G. Box, G. M. Jenkins, G. Reinsel. Analysis Time Series: Forecasting and Control. John Wiley and Sons, 2008.

Alicia Troncoso Lora et al. Electricity Market Price Forecasting Based on Weighted Nearest Neighbors Techniques. IEEE Transactions on Power Systems, Vol. 22, No. 3, pp. 1294-1301, 2007.

Francisco Martinez-Alvarez, Alicia Troncoso et al. Energy time series forecasting based on pattern sequence similarity. IEEE Transactions on Knowledge and Data Engineering, 2011.

F. Martínez-Álvarez, Alicia Troncoso Lora, José C. Riquelme, J. S. Aguilar-Ruíz. Discovery of Motifs for Forecast Outlier Occurrence in Time Series. Pattern Recognition Letters, Vol. 32, pp. 1652–1665, 2011.

S. Gelper, R. Fried, and C. Croux. Robust forecasting with exponential and holt-winters smoothing. Journal of Forecasting, 29:285–300, 2010.

F. Martínez-Álvarez, A. Troncoso, G. Asencio-Cortés, J. C. Riquelme. A Survey on Data Mining Techniques Applied To Electricity-Related Time Series Forecasting. Energies, Vol. 8, no. 11, pp. 13162-13193, 2015.

Data Streams - Flujos de datos

Charu C. Aggarwal (Ed) Data Streams: Models and Algorithms. Springer, 2007. Enlace web: http://charuaggarwal.net/streambook.pdf

Joao Gama. Knowledge Discovery from Data Streams. Chapman and Hall/CRC, 2010. Enlace web: www.liaad.up.pt/area/jgama/DataStreamsCRC.pdf

Joao Gama et al. A Survey on Concept Drift Adaptation, ACM Computing Surveys, Vol. 1, No. 1, Article 1, 2013.

Joao Gama, A survey on learning from data streams: current and future trends. Progress in Artificial Intelligence, Volume 1, Issue 1, pp 45–55, 2012.

Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. Cambridge University Press, 2013. Enlace web: http://www.mmds.org/#ver21

Albert Bifet, Geoff Holmes, Richard Kirkby and Bernhard Pfahringer. Data Streams Mining: A Practical Approach, University of Waikato, 2011.

MOA (Massive Online Analysis) http://moa.cms.waikato.ac.nz/

Multi-etiquetas

K. Dembczynski, W. Waegeman, W. Cheng, E. Hüllermeier, On label dependence and loss minimization in multi-label classification, Machine Learning 88, pp. 5–45. 2012.

K. Dembczynski, W. Cheng, E. Hüllermeier, Bayes Optimal Multilabel Classification via Probabilistic Classifier Chains, in: ICML, 2010, pp. 279–286. 2010.

E. Gibaja, S. Ventura, Multilabel learning: a review of the state of the art and on going research,Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 4, no. 6, pp. 411–444, 2014.

F. Herrera, F. Charte, A.J. Rivera, A.J., M.J. del Jesus, Multilabel Classification: Problem Analysis, Metrics and Techniques. Springer. 2016.

O. Luaces, J. Díez, J. Barranquero, J. J. del Coz, A. Bahamonde, Binary relevance efficacy for multilabel classification, Progress in Artificial Intelligence 4, pp. 303–313. 2012.

G. Madjarov, D. Kocev, D. Gjorgjevikj, S. Dzeroski, An extensive experimental comparison of methods for multi-label learning, Pattern Recognition 45, pp. 3084–3104. 2012.

E. Montañés, R. Senge, J. Barranquero, J. R. Quevedo, J. J. del Coz, and E. Hüllermeier, Dependent binary relevance models for multi-label classification, Pattern Recognition, vol. 47, no. 3, pp. 1494 – 1508, 2014.

G. Tsoumakas, I. Katakis, I. Vlahavas, Mining multi-label data, in: Data Mining and Knowledge Discovery Handbook, pp. 667-685. Springer US, 2009.

G. Tsoumakas, I. Vlahavas, Random k-Labelsets: An Ensemble Method for Multilabel Classification, in: ECML/PKDD’07, LNCS, Springer, 2007, pp. 406–417.

J. Read, B. Pfahringer, G. Holmes, E. Frank, Classifier chains for multi-label classification, Machine Learning 85, pp. 333–359. 2011.

Cuantificación

J. Barranquero, P. González, J. Díez, J. J. Del Coz, On the study of nearest neighbor algorithms for prevalence estimation in binary problems, Pattern Recognition 46 (2) pp. 472–482. 2013.

J. Barranquero, J. Díez, J. J. del Coz, Quantification-oriented learning based on reliable

classifiers, Pattern Recognition 48 (2) pp. 591–604. 2015.

A. Bella, C. Ferri, J. Hernández-Orallo, M. J. Ramirez-Quintana, Quantification via probability estimators, in: IEEE International Conference on Data Mining (ICDM’10), 737–742, 2010.

M.C. Du Plessis, M. Sugiyama, Class prior estimation from positive and unlabeled data. IEICE Transactions on Information and Systems 97, 5 pp. 1358–1362. 2014.

Andrea Esuli, Fabrizio Sebastiani. 2015. Optimizing Text Quantifiers for Multivariate Loss Functions. ACM Transactions on Knowledge Discovery Data 9, 4 27:1–27. 2015.

G. Forman, Quantifying counts and costs via classification, Data Mining and Knowledge Discovery 17 (2) pp. 164–206. 2008.

W. Gao, F. Sebastiani, Tweet Sentiment: From Classification to Quantification, in: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ACM, 97–104, 2015.

V. González-Castro, R. Alaiz-Rodríguez, E. Alegre, Class Distribution Estimation based on the Hellinger Distance, Information Sciences 218 (2013) 146–164.

D. J. Hopkins, G. King, A method of automated nonparametric content analysis for social science. American Journal of Political Science 54, 1 pp. 229–247. 2010.

L. Milli, A. Monreale, G. Rossetti, F. Giannotti, D. Pedreschi, F. Sebastiani, Quantification trees, in: IEEE International Conference on Data Mining (ICDM’13), 528–536, 2013.

J. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. Chawla, F. Herrera, A unifying view on dataset shift in classification, Pattern Recognition 45 (1) pp. 521–530. 2012.

P. Pérez-Gallego, J. R. Quevedo, J.J. del Coz, Using ensembles for problems with characterizable changes in data distribution: A case study on quantification. Information Fusion 34 pp. 87–100. 2017.

M. Saerens, P. Latinne, C. Decaestecker, Adjusting the outputs of a classifier to new a priori probabilities: A simple procedure. Neural Computation 14, 1 pp. 21–41. 2002.

M. Sugiyama, T. Kanamori, T. Suzuki, M.C. du Plessis, S. Liu, I. Takeuchi, Density-difference estimation. Neural Computation 25, 10 pp. 2734–2775. 2013.

G. I. Webb, R. Hyde, H. Cao, H. L. Nguyen, F. Petitjean, Characterizing concept drift. Data Mining and Knowledge Discovery pp. 1–31. 2015.