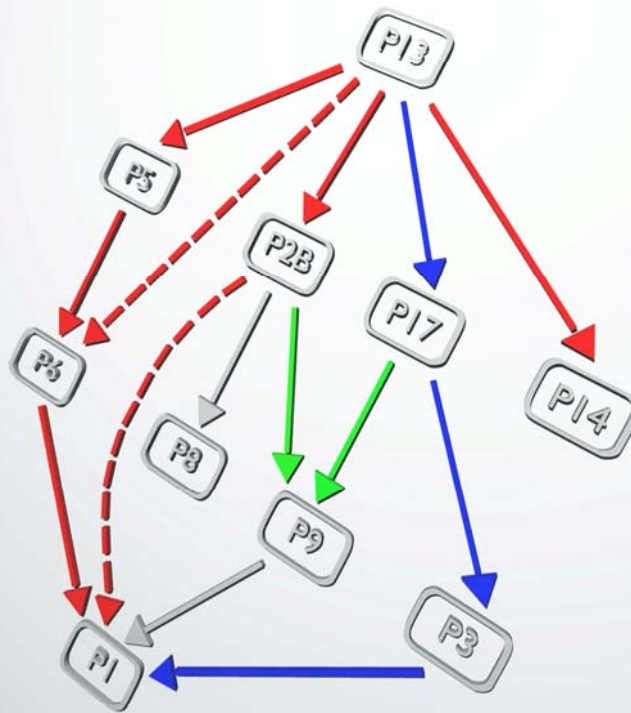


Teoría y aplicaciones del Análisis Estadístico Implicativo

Primera aproximación en lengua hispana



Eds: P. Orús, L. Zamora, P. Gregori.



Datos catalográficos

Teoría y aplicaciones del Análisis Estadístico Implicativo: primera aproximación en lengua hispana.

Editores: Pilar Orús, Larisa Zamora y Pablo Gregori

ISBN: 978-84-692-3925-4

DEP. LEGAL: CS 224-2009

Editores:

Pilar Orús Báguena
Profesora Titular de Universidad
Departamento de Matemáticas
Escuela Superior de Tecnología y Ciencias Experimentales
Universitat Jaume I de Castellón, España

Larisa Zamora Matamoros
Profesora Titular
Departamento de Matemática
Facultad de Matemática y Computación
Universidad de Oriente, Cuba

Pablo Gregori Huerta
Profesor Contratado Doctor
Departamento de Matemáticas
Escuela Superior de Tecnología y Ciencias Experimentales
Universitat Jaume I de Castellón, España

© Del texto: Los Editores, 2009

© De la presente edición: Departamento de Matemáticas, Universitat Jaume I de Castellón y Facultad de Matemática y Computación, Universidad de Oriente de Santiago de Cuba, 2009

Impresión: Innovació Digital Castelló, s.l.u.

Diseño de cubierta: Óscar Muñoz Muñoz

ISBN: 978-84-692-3925-4

Depósito legal: CS 224-2009

Presentación

El libro que presentamos es una obra colectiva, fruto de la cooperación de diversos investigadores, instituciones y países, con lenguas diferentes, pero con objetivos comunes. Entre las instituciones que han contribuido de manera decisiva a la existencia de este libro, destacaremos tres: el Grupo Internacional de Análisis Estadístico Implicativo (*Groupe International d'Analyse Statistique Implicative*, http://math.unipa.it/~grim/asi/asi_index.htm), y dos universidades, la Universidad Jaume I de Castellón (España) y la Universidad de Oriente de Santiago de Cuba (Cuba).

El Grupo Internacional de Análisis Estadístico Implicativo (ASI) decidió en 2005, en sus III Jornadas Internacionales, elegir a la Universidad Jaume I de Castellón como sede para celebrar sus siguientes Jornadas Internacionales en 2007. Con dicha decisión se pretendía continuar el objetivo de difundir estas novedosas técnicas de análisis estadístico multivariante a un país de idioma diferente a los que había utilizado habitualmente el ASI en sus publicaciones: el francés, principalmente, por el origen de sus creadores (R. Gras y colaboradores), y el inglés, por obvias razones de necesidad de comunicación científica. El italiano se había incorporado más recientemente a esta labor de difusión del Análisis Implicativo en las mencionadas III Jornadas, celebradas en Palermo, con ese mismo objetivo.

La elección de la Universidad Jaume I (UJI) permitió presentar, por primera vez en España, este método de análisis y sus técnicas a los investigadores de lengua hispana, añadiendo, además, la posibilidad de incorporar, al grupo internacional, investigadores que comparten esta lengua, de países como Cuba y Chile, y que realizan estancias de investigación frecuentes en el Departamento de Matemáticas de la UJI, muy especialmente profesores de la Universidad de Oriente (UO), con la que la Universidad Jaume I tiene establecidos fuertes vínculos de cooperación (en nuestro caso, a través de sus respectivos Departamentos de Matemáticas y de la Facultad de Matemática y Computación). Las actas de las IV Jornadas Internacionales ASI fueron ya una muestra de la cooperación entre estas instituciones y de la incipiente incorporación del español como lengua de las contribuciones presentadas en ellas.

Este libro es un paso más en este objetivo común: es la primera publicación sobre el Análisis Estadístico Implicativo realizada, íntegramente, en lengua hispana. Hemos procedido a realizar una selección y traducción de las últimas aportaciones teóricas y prácticas sobre este método de análisis, presentadas en gran parte en las últimas Jornadas realizadas hasta el momento sobre este tema, y cuyo

interés no vamos a comentar aquí, ya que el profesor Einoshin Suzuki lo hace brillantemente en el prefacio que ha escrito para nuestro libro. Queremos señalar, de forma especial, el interés de dos capítulos: el capítulo colectivo *Conceptos fundamentales del Análisis Estadístico Implicativo (ASI) y su soporte computacional CHIC*, a partir de un trabajo previo de la profesora Larisa Zamora para presentar el ASI en la UO, es una recapitulación y traducción de todos los conceptos teóricos básicos del Análisis Estadístico Implicativo, con ejemplos aclaratorios de cada uno de ellos, utilizando el programa informático *CHIC*, que implementa y grafica los cálculos de las diferentes técnicas del ASI: árboles jerárquicos de similaridad, grafos de implicación y árboles jerárquicos de cohesión. El capítulo del profesor Einoshin Suzuki, *Descubriendo Reglas de Excepción Estructuradas*, basado en una de sus contribuciones al congreso *15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, nos ha sido gentilmente proporcionado para su traducción e inclusión en el libro.

La estructura del libro, articulado en tres partes, está marcada por la naturaleza de los capítulos: las contribuciones teóricas al desarrollo del ASI las encontraremos en la primera parte, y las aplicaciones de este método y sus técnicas a diversos temas y ámbitos aparecerán en la segunda y tercera parte. Si bien, la mayoría de las aplicaciones están realizadas en el ámbito de la didáctica de las Matemáticas (origen del ASI), hemos querido señalar que en la actualidad ya existen aplicaciones en otros campos, diferenciando una tercera parte con los capítulos seleccionados por el interés especial que supone mostrar la utilización del ASI como instrumento metodológico en nuevos ámbitos de investigación científica y social.

Queremos agradecer a todos los autores su disponibilidad y diligencia para proporcionarnos versiones en lengua hispana de sus capítulos, o bien su permiso para poder disponer de los capítulos en versión original, junto con su confianza, para la realización de las traducciones. Agradecemos así mismo a los traductores su excelente, a la vez que desinteresado, trabajo de traducción; por ello hemos querido incluir también su nombre, cuando este trabajo no ha sido hecho por los propios autores. De manera especial queremos agradecer el gran volumen de trabajo de traducción de los compañeros del Departamento de Matemáticas de la Universidad de Oriente y de la coordinación ejercida por su directora, la profesora Zamora. Así mismo queremos agradecer su gran contribución al trabajo de edición, realizado principalmente durante su estancia de investigación en la Universidad Jaume I (financiada por las ayudas Bancaja-UJI 2008), en colaboración con el profesor Pablo Gregori, del Departamento de Matemáticas de la UJI. Gracias a todos.

Por último, y de manera especial, queremos agradecer a Régis Gras, profesor emérito de la Université de Nantes, la confianza y apoyo, en todo momento, a nuestro proyecto, y a Einoshin Suzuki, profesor de la Kyushu University, por la cesión de su artículo y la redacción de nuestro prefacio.

Este libro, en habla hispana, junto con *Statistical Implicative Analysis: theory and applications* (Springer, 2007), y el recientemente aparecido *Analyse Statistique Implicative; une méthode d'analyse de données pour la recherche de causalités* (Cépadues, 2009), cumple el objetivo común de ampliar la difusión del Análisis Estadístico Implicativo a un mayor número de investigadores y utilizadores potenciales en diversos ámbitos, a nivel mundial.

Castellón, 2 de Julio de 2009.

Pilar Orús
Directora del Dpto de Matemáticas
Universidad Jaime I. Castellón.
España

Larisa Zamora
Directora del Dpto de Matemáticas
Universidad de Oriente. Santiago de Cuba.
Cuba

Prefacio (traducción)

El descubrimiento de reglas está reconocido como un campo representativo de la minería de datos, y los criterios de evaluación, que distinguen las reglas "interesantes" de las demás desempeñan un papel crucial en el descubrimiento de reglas. El trabajo de Régis Gras en el campo del Análisis Estadístico Implicativo (ASI) se remonta a la década de los 70 y es ampliamente reconocido por haber realizado contribuciones significativas a la investigación sobre criterios de evaluación. Este libro contiene, en español, una gran parte de los trabajos presentados en ASI4 (4èmes Rencontres Internationales Analyse Statistique Implicative), siendo ésta la última de la serie de conferencias internacionales que marcan la orientación del ASI, que se celebró del 18 al 21 de octubre de 2007 en el Departamento de Matemáticas de la Universitat Jaume I de Castellón en España. Este libro representa la primera aproximación a una colección completa y detallada de artículos sobre avances recientes del ASI en términos de metodologías y aplicaciones, en un idioma utilizado diariamente por unos 350 millones de personas en todo el mundo.

La intensidad de la implicación, el primer método del ASI en su forma inicial, evalúa el grado de interés de una regla $Y \rightarrow X$ en términos de la cantidad $n(\bar{X} \wedge Y)$ de sus contraejemplos, cada uno de los cuales cumple $\bar{X} \wedge Y$. Es una medida de interés basada en la probabilidad de obtener una regla cuyo número de contraejemplos es a lo sumo $n(\bar{X} \wedge Y)$ de los datos disponibles, y ha implicado a un gran número de investigadores y personas de negocios debido a su fundamentación teórica y méritos prácticos. CHIC (Cohesive Hierarchical Implicative Classification) es un programa informático equipado con una interfaz gráfica que permite una fácil utilización de diversas técnicas del ASI para una amplia gama de usuarios, desde expertos en el análisis de datos hasta los profesionales con poca base informática.

El libro contiene explicaciones sobre el ASI y CHIC, así como ejemplos de sus aplicaciones, especialmente en educación, didáctica, y psicología, pero también en bioinformática e ingeniería del conocimiento. Algunas extensiones del ASI en el contexto general de la minería de datos también se presentan.

Este libro será de interés para los creadores de sistemas de minería de datos, así como investigadores, estudiantes y profesionales dedicados a la minería de datos y análisis estadístico de los datos.

Profesor Einoshin Suzuki
Departamento de Informática,
Escuela Superior de
Ciencias de la Información
e Ingeniería Eléctrica,
Universidad de Kyushu, Japón

Preface (original)

Rule discovery is recognized as a representative subject of data mining, and evaluation criteria, which distinguish “interesting” rules from others play a crucial role in rule discovery. Régis Gras’s work on Statistical Implicative Analysis (SIA) dates back to the 1970’s and is widely acknowledged to have made significant contributions to the research on evaluation criteria. This book contains a large part of ASI4 (4èmes Rencontres Internationales Analyse Statistique Implicative), a leading international conference in SIA held during 18-21 October 2007 at Departamento de Matemáticas Universitat Jaume I, Castellón, España, with some new insights, in Spanish language. This book represents the first approach of a collection of comprehensive and detailed reports on recent progress of SIA in terms of methodologies and applications, in a language used daily by approximately 350 million of people world-wide.

Intensity of implication, the first method in SIA in its initial form, evaluates the degree of the interestingness of a rule $Y \rightarrow X$ in terms of the number $n(\bar{X} \wedge Y)$ of its counterexamples each of which satisfies $\bar{X} \wedge Y$. It is an interestingness measure based on the probability of obtaining a rule of which number of counterexamples is at most $n(\bar{X} \wedge Y)$ from the given data and has been involving a large number of academic researchers and business persons due to its theoretical beauty and practical merits. CHIC (Cohesive Hierarchical Implicative Classification), a software equipped with a graphical interface, allows easy use of various techniques in SIA for a wide range of users from experts in data analysis to practitioners with little background on computer science.

The book contains explanations of SIA and CHIC as well as examples of their applications especially in education, didactics, and psychology; but also bioinformatics and knowledge engineering. Extensions of SIA in the context of general data mining are also presented.

This book should be of interest to developers of data mining systems as well as researchers, students, and practitioners devoted to data mining and statistical data analysis.

Professor Einoshin Suzuki
Department of Informatics,
Graduate School of
Information Science
and Electrical Engineering,
Kyushu University, Japan

Índice

Presentación	iii
Prefacio(traducción)	vii
Preface (original)	ix
Índice	xi

Parte I: Trabajos teóricos

1. El Análisis Estadístico Implicativo (ASI) en respuesta a problemas que le dieron origen	3
2. CHIC: utilización y funcionalidades	51
3. Conceptos fundamentales del Análisis Estadístico Implicativo (ASI) y su soporte computacional CHIC.....	65
4. Descubriendo Reglas de Excepción Estructuradas.....	103
5. Regla y R-regla de excepción en Análisis Estadístico Implicativo	127
6. Una medida comparativa en didáctica de las matemáticas entre el análisis a priori y la contingencia	143
7. Reflexiones sobre el análisis a priori de los cuestionarios basadas en técnicas del Análisis Estadístico Implicativo.....	159

Parte II: Aplicación del ASI a la Didáctica de las Matemáticas

8. Previsión y comunicación de valores a través del gráfico.....	193
9. La comprensión de la divisibilidad en \mathbb{N} . Un análisis implicativo.....	215
10. Aplicación de CHIC al estudio de las funciones elementales	235
11. Una investigación sobre la formalización que hacen los alumnos de un problema de la teoría de números	265

12. Tratamiento de datos, lógica y estadística elemental en la enseñanza preuniversitaria. Un estudio mediante la aplicación del CHIC.....	279
13. ¿Cómo enseñan las derivadas los profesores de cálculo, en la Universidad?.....	299
14. ¿Cómo se enseña en la Universidad? El caso de los Números Reales	329
15. El análisis implicativo como medio de evaluación de los estudiantes futuros maestros de escuela y de su formación en geometría plana	353
16. Criterios de adopción y utilización del libro didáctico de matemática en la enseñanza fundamental del nordeste brasileño.	375
Parte III: Aplicaciones del ASI en otros ámbitos	
17. “Representaciones” de profesores ordinarios sobre situaciones de corrección y sobre la planificación del curriculum a propósito de la función inversa en la transición liceo/Universidad en Francia.....	403
18. Análisis estadístico implicativo, análisis textual. Un ejemplo de complementariedad.....	427
19. Enfoque bayesiano “oculto” y enfoque frecuentista “ambiguo” en los manuales franceses de Première S y ES.....	447
20. Un método implicativo para el análisis de datos de expresión de genes	461
21. Relaciones entre tipos de violaciones de libertades sindicales garantizadas por los convenios de la OIT: Un análisis estadístico implicativo de los resultados de un análisis automático de textos	483