

Estudio de relaciones causales entre indicadores de promoción del primer año de la carrera Ciencia de la Computación con el análisis implicativo modal

Study of causal relationships among indicators of academic performance in the freshman year of Computer Science by using modal implicative analysis

Dra. C. Larisa Zamora-Matamoros, larisa@uo.edu.cu;

Dr. C. Jorge Díaz-Silvera, jdiaz@uo.edu.cu

Universidad de Oriente, Santiago de Cuba, Cuba

Resumen

El análisis estadístico implicativo es un método de análisis de datos no simétrico cuyo objetivo principal contempla la estructuración de datos, interrelacionando sujetos y variables, la extracción de reglas inductivas entre las variables y a partir de la contingencia de éstas, la explicación y en consecuencia una determinada previsión en distintas ramas del saber. Lo integran dos técnicas de análisis de datos, el análisis cohesivo y el implicativo, además del clasificatorio o de similaridad. El objetivo del presente trabajo de investigación es mostrar la aplicabilidad del análisis estadístico implicativo para revelar posibles relaciones de similaridad, propensión y cohesión entre el rendimiento académico de estudiantes que ingresan a la carrera de Ciencia de la Computación y el rendimiento que muestran en las asignaturas de corte matemático y de Programación que reciben en el primer año de la mencionada carrera. Los datos recopilados fueron procesados usando el software SIASI para datos modales.

Palabras clave: análisis estadístico implicativo, SIASI, variables modales, rendimiento académico.

Abstract

The Statistical Implicative Analysis (SIA) is a method of non-symmetrical analysis of data whose main objective is the structuring of data, interrelating individuals and variables, the extraction of inductive rules among the variables and from their contingency, the explanation and in consequence a certain prediction in different knowledge branches. The SIA holds two techniques of analysis of data, the cohesive analysis and the implicative analysis, along with the classificatory or similarity analysis. The objective of the present research is to reveal possible similarity, propensity and cohesion relationships among the academic results of students coming from high schools that enter to Computer Science career and the results that they show in undergraduate courses related to Mathematics and Programming, which they receive in the first year of the mentioned career. The gathered data were processed using the software SIASI for modal data.

Key words: statistical implicative analysis, SIASI, modal variables, academic performance.

Introducción

El Análisis Estadístico Implicativo (AEI) es un método de análisis de datos no simétrico desarrollado por Régis Gras y sus colaboradores a partir de la década de los 70, cuyo objetivo principal contempla la estructuración de datos interrelacionando sujetos y variables, la extracción de reglas inductivas entre las variables y a partir de la contingencia de estas reglas, la explicación y en consecuencia una determinada previsión en distintos ámbitos del saber, como la psicología, sociología, biología, etc.(Zamora, Gregori & Orús, 2009).

El origen de este método es la modelización estadística de la cuasi-implicación: cuando la variable o la conjunción de variables a es observada en la población, entonces generalmente la variable b lo es también. Su objetivo inicial fue responder a la pregunta: si un objeto posee cierta propiedad, entonces ¿posee alguna otra? La respuesta puede no ser completamente afirmativa, sin embargo es posible constatar que aparecen ciertas tendencias. El AEI tiene por objetivo constatar tales tendencias en un conjunto de propiedades.

El AEI está integrado por dos técnicas de análisis de datos, el análisis cohesitivo y el análisis implicativo, las cuales se encuentran implementadas en el sistema informático CHIC (*Classification Hiérarchique Implicative et Cohésitive*), el cual fue concebido por Régis Gras y ha sido desarrollado, en sus versiones más recientes, por Couturier (2008). Este software incorpora, además de las dos técnicas antes mencionadas, la clasificación mediante el análisis de similaridad.

El programa informático CHIC brinda variadas posibilidades, pero a la vez presenta limitaciones analizadas por Portuondo (2014). Desde el 2011 en la Universidad de Oriente, en Santiago de Cuba, se ha venido trabajando en la elaboración del sistema informático para el análisis estadístico implicativo, el cual se ha denominado abreviadamente SIASI, para implementar las técnicas del AEI y superar las desventajas que hasta el momento presenta el CHIC.

Dos de los indicadores cuantitativos del rendimiento como referentes de calidad y eficiencia de un programa de formación profesional o de un curso de una materia específica son la calificación o nota por materia y para el ciclo de formación(Torres *et al*, 2004).

Teniendo en cuenta los indicadores antes mencionados, en el presente trabajo se muestra la aplicación del AEI para revelar posibles relaciones de similitud, propensión y cohesión entre el índice académico en la enseñanza preuniversitaria de estudiantes que ingresan a la carrera de Ciencia de la Computación y las notas que muestran en las asignaturas de corte matemático y de Programación que reciben en el primer año de esta carrera, usando variables modales, a diferencia del análisis realizado por Zamora y Díaz (2008) en el cual se emplearon variables binarias.

Desarrollo

Los datos para realizar la investigación fueron tomados de los registros de 18 estudiantes de la carrera de Ciencia de la Computación al concluir el primer año en el curso 2015-16, ubicados en la secretaría docente de la facultad de Ciencias Naturales y Exactas.

Las variables objeto de estudio fueron: el índice académico de preuniversitario (IA), la nota de Matemática en el examen de ingreso a la educación superior (MatEI), y las notas obtenidas en Lógica (Log), Geometría (Geo), Álgebra I (Alg_I) (todas estas asignaturas del primer semestre), Análisis Matemático I (AM_I), Álgebra II (Alg_II) (asignaturas del segundo semestre), Programación (Prog) (asignatura año) y promovido (Prom), que representa el promedio de notas de las asignaturas de perfil matemático y de Programación con que promueve a segundo año.

En el estudio se consideraron algunas variables suplementarias. Una variable suplementaria es una variable intrínseca, por ejemplo un descriptor (edad, sexo, categoría socio - profesional, etc.), que no interviene en la formación de las reglas o clases de reglas. Sin embargo, podemos estar interesados en investigar el papel que ésta desempeña en la formación de las reglas o clases obtenidas. Por ejemplo, si la variable suplementaria es el sexo, podemos estar interesados en determinar qué sexo es el que ha aportado más a la obtención de esas reglas o clases de reglas, para lo cual se definen la tipicidad y contribución de esas variables suplementarias.

En el trabajo se consideraron como variables suplementarias el sexo (Sexo1: masculino y Sexo0: femenino), la procedencia (PE1: IPU, PE3: IPVCE), la provincia de residencia (Prov acompañado del código de la provincia), la convocatoria en que alcanzó la plaza (Conv1: ordinaria y Conv2: extraordinaria), opción en que solicitó la carrera (Opc seguido del número de opción en que solicitó la carrera), la fuente de ingreso (FI1:

examen de ingreso, FI3: concurso, FI7: olimpiada yFI9: cadete) y la clase de estudiante (CE1: becado yCE3: externo).

Dado que la mayoría de las variables investigadas son modales (notas de los estudiantes según la clasificación de la educación superior cubana de 2 a 5), se aplica el análisis estadístico implicativo para variables modales (Lagrange, 1998), con 4 modalidades. Para el caso del índice académico y la nota del examen de ingreso de Matemática, se emplea la codificación siguiente, en la cual la x representa la nota del estudiante: 5 para $90 \leq x \leq 100$, 4 para $80 \leq x < 90$, 3 para $60 \leq x < 80$ y 2 para $x < 60$.

En el caso de la variable Prom, para estudiantes que promueven al segundo año, la codificación empleada fue la siguiente: 5 para promedios en el intervalo $[4,5; 5]$, 4 para promedios en $[3,5; 4,5)$, 3 para promedios menores o iguales a 3,5. Se reserva la modalidad 2 para los estudiantes que no promovieron. Hay que tener en cuenta que si el estudiante no aprueba Programación, no puede pasar al segundo año.

En la tabla 1 se muestran la media y la desviación estándar de cada una de las variables principales consideradas en la investigación. Se observa un cambio brusco entre los resultados obtenidos por los estudiantes en la enseñanza media superior, representados por las variables índice académico de preuniversitario (IA) y la nota de Matemática en el examen de ingreso a la educación superior (MatEI), y las notas de las asignaturas estudiadas en el primer año de la carrera.

Tabla 1. Media y desviación estándar de las calificaciones

Variable	Media	Desviación Estándar
<i>IA</i>	4,94	0,23
<i>MatEI</i>	3,83	0,96
Log	3,78	1,13
Geo	3,11	1,24
Alg_I	3,00	1,29
Alg_II	3,06	1,18
Prog	3,50	1,30
AM_I	3,11	1,33
Prom	3,11	1,29

A partir de los datos recopilados se confeccionó una matriz de datos, la cual se usó para la corrida con el sistema SIASI.

Antes de aplicar las técnicas del ASI se realizó el análisis de similaridad o clasificatorio, el cual es una técnica exploratoria diseñada para revelar agrupaciones naturales dentro de una colección de datos. La similaridad es la medida de semejanza usada para agrupar los objetos. El árbol de similaridad obtenido se muestra en la figura 1.

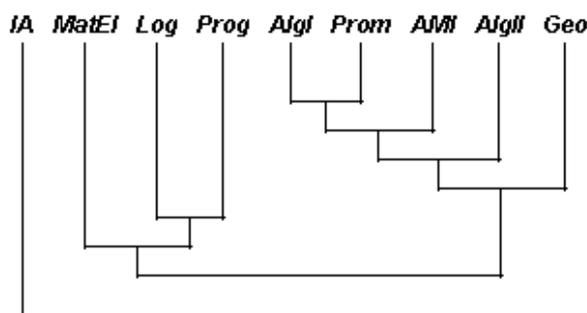


Figura 1. Árbol de similaridad de las calificaciones de estudiantes

Se aprecia que se forman dos grupos de asignaturas similares en cuanto a los resultados obtenidos por los estudiantes en las mismas:

Grupo 1 (G1): {AlgI, Prom, AMI, AlgII, Geo} y

Grupo 2 (G2): {MatEI, Log, Prog}

En G1 quedan agrupadas la mayoría de las notas alcanzadas en el primer año de la carrera, exceptuando Lógica y Programación, con un índice de similaridad de 0,9312. En G2 se agrupan las notas de Matemática en el examen de ingreso y las de Lógica y Programación del primer año, con una similaridad de 0,6858.

Asímismo se observa en el árbol de similaridad un variable que no se asemeja a ninguna otra, que es el índice académico de los estudiantes en el preuniversitario, lo cual muestra que el índice acumulado por los estudiantes en su tránsito por la educación media superior no guarda ninguna semejanza con los resultados que obtienen en el primer año de la carrera, ni con la nota en el examen de ingreso de Matemática.

El grafo de la figura 2 muestra las relaciones de propensión entre las notas, las cuales se interpretan en el siguiente sentido: hay propensión de una nota a hacia otra b , si como promedio se encuentran pocos estudiantes para los cuales a es fuerte y b es débil (o sea, que a tome una modalidad mayor que b).

En el grafo solo se han representado las relaciones en tres intervalos de índices de propensión expresados en porcentaje: en rojo [96, 100], en azul [93, 96) y verde [90, 93).

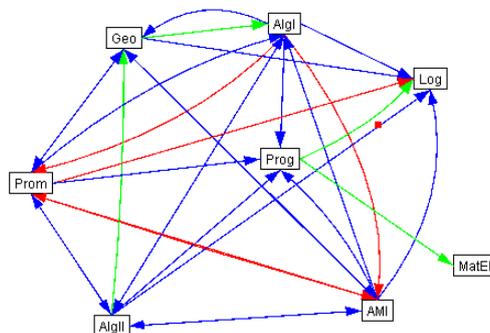


Figura 2. Grafo de propensión de las calificaciones de estudiantes

En la figura 2 puede observarse, para índices de propensión superiores al 90%, que:

- No aparece relación directa de ninguna nota contemplada en la carrera con el índice académico de la enseñanza media superior.
- Hay influencia directa más fuerte al promovido desde las asignaturas de Álgebra I y Análisis Matemático I y en menor proporción desde Álgebra II y Geometría.
- Hay influencia directa del Análisis Matemático I, el Álgebra I y Álgebra II hacia Programación, con índices de propensión del 94%, 95% y 94% respectivamente; en cambio, no se nota influencia de Programación hacia las notas de las asignaturas de corte matemático, salvo la de Lógica.
- Cabe destacar la doble implicación entre las variables Prom y las de corte matemático, exceptuando Lógica, lo cual significa que hay pocos estudiantes con diferencias entre la modalidad alcanzada en la variable promovido y la nota de la asignatura en particular. Estas diferencias se incrementan a medida que disminuye el índice de propensión.
- La relación de Prom a Prog indica, con una intensidad de propensión del 95%, que existen pocos estudiantes que no habiendo promovido hayan aprobado Programación. Dicho de otra forma, la modalidad (nota) que alcanza un estudiante en la variable Prog es como mínimo la modalidad que alcanzó en Prom con la intensidad planteada.

- Álgebra I es una asignatura dominante en cuanto a la calidad de las notas que alcanzan los estudiantes del primer año, en el sentido de que en el resto de las asignaturas las notas alcanzadas por los estudiantes serán como mínimo las alcanzadas en Álgebra I con pocas excepciones.

En la figura 3 se muestra el árbol cohesitivo, el cual permite la obtención de un conjunto de reglas y meta reglas que expresan relaciones de causalidad entre las habilidades adquiridas por los estudiantes en las materias tratadas en el presente trabajo. La cohesión surge como una medida de la calidad de la propensión de una regla del tipo $R: a \rightarrow b$.

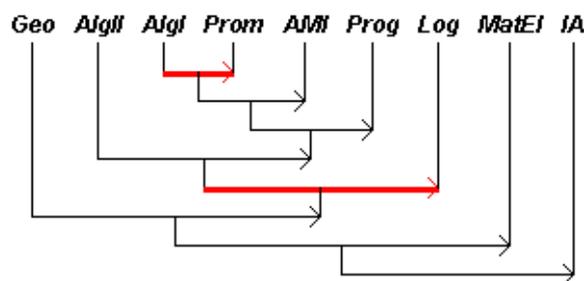


Figura 3. Árbol cohesitivo de calificaciones de estudiantes

A partir del árbol cohesitivo se desprende la formación de las siguientes reglas y meta reglas, que permiten dar cuenta de las estructuras que los estudiantes han construido durante el aprendizaje en su primer año en la carrera, lo cual se muestra en la tabla 2.

Tabla 2. Formación de reglas y meta reglas

Regla	Descripción	Índice de cohesión
R ₁	AlgI ⇒ Prom	0,9773
R ₂	R ₁ ⇒ AMI	0,9756
R ₃	R ₂ ⇒ Prog	0,9664
R ₄	AlgII ⇒ R ₃	0,9604
R ₅	R ₄ ⇒ Log	0,9538
R ₆	Geo ⇒ R ₅	0,9420
R ₇	R ₆ ⇒ MatEI	0,9068
R ₈	R ₇ ⇒ IA	0,8145

Por cuestiones de espacio solo se presentan las interpretaciones de la regla R₁ y la meta regla R₂. El resto de las interpretaciones de hacen de forma similar.

- R_1 : la calificación promedio que recibe el estudiante en el año es como mínimo la que recibió en Álgebra I, por lo que se puede decir que la nota en esta asignatura es la que define la nota promedio del año. En términos de habilidades se puede afirmar, con un índice de cohesión de 0,9773, que las habilidades que como promedio alcanza el estudiante en el curso académico son mayores que las alcanzadas en Álgebra I y es poco probable que un estudiante que no haya promovido haya aprobado Álgebra I.
- R_2 : la nota en Análisis Matemático I es mayor o igual que la menor de las notas obtenidas en la asignatura Álgebra I y en promovido. En otras palabras, las habilidades alcanzadas por el estudiante en Análisis Matemático I serán al menos iguales a las alcanzadas en Algebra I y las generales que desarrolló en el año. Por otra parte se puede afirmar que es poco probable que un estudiante que desaprobe el Análisis Matemático I desaprobe el Álgebra I o que promueva.
- R_3 : esta regla pone de manifiesto que es poco probable, con una cohesión de 0,9664, que un estudiante obtenga en Programación una nota inferior a la nota mínima alcanzada en las asignaturas de Álgebra I, Análisis Matemático I y en la variable promovido.

Algunas de las asignaturas que aparecen involucradas en las reglas y meta reglas antes enunciadas, estaban próximas en el árbol de similaridad, con excepción del índice académico en el preuniversitario. Como planteamos anteriormente, las habilidades con las que los estudiantes ingresan a la carrera de Ciencia de la Computación, expresadas a través del índice académico de preuniversitario (IA), no se asemejan a los niveles de habilidades que alcanzan en el primer año de dicha carrera. Por otra parte, el árbol cohesivo muestra que dichas habilidades están por debajo de las habilidades adquiridas en el primer año de la carrera.

En la figura 4 se muestran, a modo de ejemplo de salida del SIASI, los valores de riesgo de la contribución de cada una de las variables suplementarias definidas, junto con la cantidad de individuos presentes en el grupo óptimo para la regla $(AlgI \Rightarrow Prom) \Rightarrow AMI$ obtenida del árbol cohesivo. Los resultados obtenidos del análisis de la typicalidad de las variables suplementarias no se incluyen debido a que fueron idénticos a los obtenidos en el análisis de la contribución. De esta figura podemos resumir que las variables suplementarias más típicas y contributivas a la formación de la regla son: el sexo femenino (Sexo0), provenientes de IPVCE (Pre9), de la provincia de Holguín

(Prov11), presentados en la convocatoria extraordinaria (Conv2), los estudiantes que solicitaron la carrera en segunda, tercera, cuarta y séptima opción (Opc2, Opc3, Opc4 y Opc7), estudiantes becados (CE1) y fuentes de ingreso de concurso, olimpiada y cadete (FI3, FI7 y FI9).

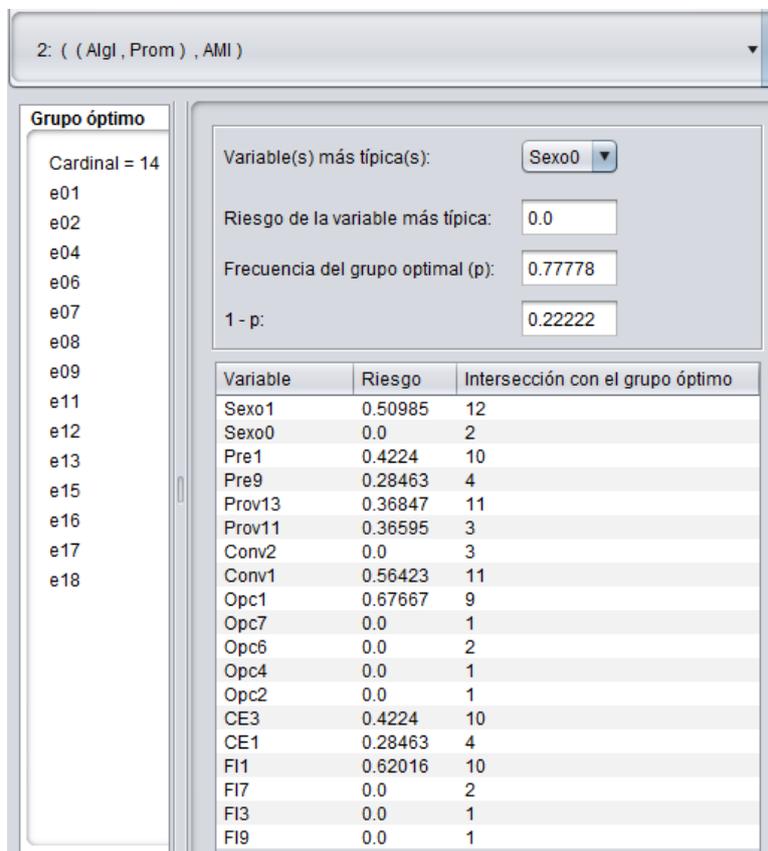


Figura 4. Tipicalidad/contribución de las variables suplementarias

Conclusiones

- 1. En el presente trabajo se ha aplicado el análisis estadístico implicativo a un estudio sobre el rendimiento académico de los estudiantes del primer año de la carrera de Licenciatura en Ciencia de la Computación en el curso 2015-2016, a través del índice académico que traen de la enseñanza media superior y las notas finales en las asignaturas de corte matemático y Programación del año.*
- 2. El análisis de similaridad o clasificatorio evidencia que hay una mayor similitud en cuanto a las notas alcanzadas por los estudiantes en las asignaturas Álgebra I, Geometría, Álgebra II y Análisis Matemático I y que no hay similaridad entre el índice académico que acumularon los estudiantes en el preuniversitario con las notas en las asignaturas de corte matemático y computacional del primer año.*

3. *El análisis implicativo, expresado a través del grafo de propensión, muestra que las asignaturas de corte matemático exceptuando Lógica tienen la mayor influencia hacia el promovido, Álgebra I y Análisis Matemático I con mayor influencia. También evidencia la no existencia de influencia significativa de las notas en el examen de ingreso de Matemática hacia las notas en ninguna de las asignaturas. Estos resultados son corroborados luego por el análisis cohesivo.*
4. *El presente estudio, empleando por primera vez variables modales y como software de apoyo el SIASI, revela la existencia de una contradicción entre la supuesta preparación que traen los estudiantes de la Educación Media Superior, reflejada en los altos valores del índice académico, y los bajos resultados que obtienen en las asignaturas de corte matemático y de Programación en la carrera, que corroboran resultados expuestos por Zamora y Díaz(2008) y otros estudios posteriores, pero a un mayor nivel de precisión y valoración de la calidad de la promoción, dado que en este estudio, a diferencia de los anteriores, se emplean variables modales y no binarias. La aplicación del AEI ha permitido evidenciar relaciones que se establecen entre los resultados obtenidos en las asignaturas seleccionadas para el estudio y el resultado general del año.*

Referencias bibliográficas

1. Couturier, R. (2008). Cohesive hierarchical implicative classification. En Gras, R.; Suzuki, E.; Guillet, F.; Spagnolo, F. (Eds.). (2008). *Studies in Computational Intelligence 127* (pp. 41-53). Berlin: Springer-Verlag.
2. Lagrange, J. B. (1998). Analyse implicative d'un ensemble de variables numériques; application au traitement d'un questionnaire à réponses modales ordonnées. *Revue de Statistique Appliquée*, 46(1), 71-93.
3. Portuondo, L. (2014). Técnicas de análisis estadístico implicativo y clasificatorio para variables modales y su implementación computacional. (Tesis de maestría). Universidad de Oriente, Santiago de Cuba, Cuba.
4. Torres, A.; Garcia, V.; Cruz, O.; Ruiz, E. (2004). *La calidad de la formación del profesional universitario, su caracterización y evaluación*. UNAH. Recuperado de <http://intranet.isch.edu.cu>
5. Zamora, L.; Diaz, J. (2008). Aplicación del análisis estadístico implicativo al estudio del rendimiento académico de estudiantes de primer año de las carreras de Matemática y Ciencia de la Computación. *CADERNOS DO IME – Série Estatística*, 25, 1-16.
6. Zamora, L.; Gregori, P.; Orús, P. (2009). Conceptos fundamentales del análisis estadístico implicativo (ASI) y su soporte computacional CHIC. En Orús, P.; Zamora, L.; Gregori, P. (Eds.), *Teoría y aplicaciones del análisis estadístico implicativo: primera aproximación en lengua hispana* (págs. 65–101). Castellón de la Plana: Innovació Digital Castelló, s.l.u.