INFERENCIA PROBABILÍSTICA, ALGORITMO ID3.

Gabriel Aguirre

UTN-FRM/gabriel.aguirre@docentes.frm.utn.edu.ar
UAI/GabrielVictor.Aguirre@alumnos.uai.edu.ar
/GVAguirre72@yahoo.com.ar

Resumen: La complejidad del mundo aporta incertidumbre, limitando la posibilidad de realizar inferencias racionales. Es aquí donde la IA puede brindar soluciones computacionales a partir de métodos Inferenciales automatizados como los llamados Árboles de decisión, que pueden ser implementados como un componente fundamental en los sistemas que piensan racionalmente. Uno de ellos vincula las nociones de Probabilidad y Entropía en el marco de la Teoría de la Información; estamos hablando del clásico Iterative Dichotomiser 3 (ID3), desarrollado en 1979 por John Ross Quinlan para implementar el razonamiento basado en casos. Esta investigación se centra en el estudio de los aspectos teóricos que fundan la idea del algoritmo, sin escatimar en exaltar rasgos históricos relacionados con este proceso. Implementando computacionalmente el mismo, para proceder a la validación frente a otros productos. Explorar otras alternativas teóricas que retoman el enfoque clásico de física estadística, para implementar el algoritmo.

Palabras claves: ID3**,** Entropía, Inferencia, Incertidumbre, Árboles de Decisión.

INTRODUCCIÓN

ID3 (Iterative Dichotomiser Three) o como se lo suele traducir también aludiendo a un juego de palabras Inferential Decision Tree, es un algoritmo de clasificación y decisión basado en el principio físico de la Entropía, concebido y adaptado al ámbito de la Informática. Y como tal permite la construcción de un árbol de decisión que trata de limitar la probabilidad de error al tomar una decisión por sí o por no. Dicho algoritmo fue propuesto por el Ingeniero en Sistemas Australiano John Ross Quinlan en 1979.

Para la presentación de este trabajo de Investigación, junto con algunos de los resultados obtenidos de un Mapeo Sistemático de la Literatura hemos optado por dividirlo en cuatro tramos. Un apartado de **CONTEXTO** de la temática, a continuación, los aspectos referentes a **FUNDAMENTOS LÓGICOS,** continuamos con los **FUNDAMENTOS**

FÍSICO-MATEMÁTICO que rigen la heurística del algoritmo. Dejando para el tramo final los aspectos computacionales en **DESCRIPCIÓN GENERAL DEL ALGORÍTMO ID3** e **IMPLEMENTACIÓN COMPUTACIONAL**.

CONTEXTO

¿Qué es la inteligencia?

Russell (1983) se pregunta cómo es que los seres humanos, cuyos contactos con el mundo son breves, personales y limitados, logran, sin embargo, conocer tanto como conocen. La inteligencia puede ser definida en la literatura de diversas maneras,

> "**inteligencia**. (Del lat. intelligentia). f. 1. Capacidad de entender o comprender. 2. Capacidad de resolver problemas. 3. Conocimiento, comprensión, acto de entender. 4. Sentido en que puede tomar una proposición, un dicho o una expresión…" (Real Academia Española, 2024)

Pero todas ellas deben ser tomadas como definiciones provisionales, acuerdos convencionales, puesto que es muy difícil caracterizar la inteligencia. En palabras del biólogo Stephen Jay Gould

> "…hemos llegado a ser, en virtud de un glorioso accidente evolutivo llamado **inteligencia**, los administradores de la continuidad de la vida en la Tierra. No pedimos que se nos asignara ese papel, pero no podemos rechazarlo. Quizá no seamos los más adecuados para desempeñarlo, pero aquí

> estamos… (citado en Audesirk, Audersirk y Byers, 2008, p. 353)”

Incluso durante décadas, el hombre ha descripto la inteligencia desde su propia perspectiva, pero de hecho la inteligencia puede ser observada en los organismos en general, y por tanto no es un rasgo exclusivo del género homo. En este sentido la ciencia comienza a observar y estudiar este rasgo, no solamente en el hombre.

Una comprensión de la inteligencia no puede pretender ser sustentable, enfocándose solo en el hombre, debe tener una mirada integral y observar también el resto de los casos presentes en la naturaleza.

¿Qué es la inteligencia Artificial?

Lo dicho anteriormente nos pone ante un dilema, pero el hombre mismo es parte de este dilema. Lo que nos motiva es comprender ser curiosos, hacer preguntas. Desde el momento en que el hombre pensó en construir una máquina no orgánica para auxiliarlo es sus tareas, comenzó a develar el dilema.

Actualmente nos referimos a Sistemas Inteligente (IS), en este sentido el Institute of Electrical and Electronics Engineers (IEEE) nos dice

> **Intelligent Systems (IS)**
>
> Artificial intelligence (AI) is the study of solutions for problems that are difficult or impractical to solve with traditional methods. It is used pervasively in support of everyday applications such as email, word-processing and search, as well as in the **design and analysis of autonomous agents that perceive their environment and interact rationally with the environment.** (ACM & IEEE, 2013, p. 121)

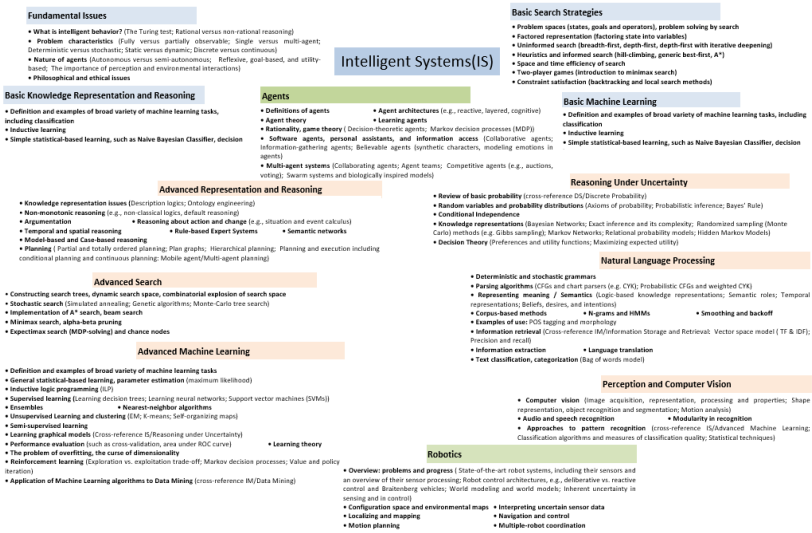
Usted está aquí.
Como parte de este estudio se realizó un ***mapeo sistemático de la literatura***. En él se plantea una primera pregunta de investigación

> **P1: ¿Qué taxonomía del Campo de la Inteligencia Artificial podría ser referenciada?**

A través de la comparación de distintos documentos y según los criterios de búsqueda y análisis establecidos se obtiene el siguiente resultado presentado en la Figura 1

Figura 1

Taxonomía del campo de la Inteligencia Artificial.



Nota. El gráfico presentado está inspirado en los resultados obtenidos en The Joint Task Force on Computing Curricula, Association for Computing Machinery (ACM) y IEEE Computer Society para el Computer Science Curricula 2013. Realizando una adecuación didáctica en "Aguirre, Gabriel Víctor TP 3.3 Mapeo Sistemático de la Literatura V1.docx", bajo las exigencias del plan de estudios para el Doctorado en Informática de la Universidad Abierta Interamericana, y enmarcada en el espacio "Metodología de la Investigación Científica" a cargo del Dr. Carlos Neil.

La taxonomía de la Inteligencia Artificial (IA) presentada en la **Figura 1** referencia de manera central la denominación actualizada del campo como **Intelligent Systems** (IS). En celeste se destacan las nociones introductorias o básicas para el abordaje del campo: **Fundamental Issues** (Cuestiones fundamentales), **Basic Knowledge Representation and Reasoning** (Representación y razonamiento de conocimientos básicos), **Basic Search Strategies** (Estrategias de búsqueda básicas) y **Basic Machine Learning** (Aprendizaje automático básico). **Agents**, Teoría de Agentes es troncal en el campo de IA y se exalta en color verde oliva en posición central.

Los sub campos que abordan temáticas en profundidad se muestran en color naranja suave - la disposición no es casual, organizadas desde categorías más específicas a otras más amplias -, podemos observar las siguientes: **Advanced Representation and Reasoning** (Representación y razonamiento avanzados), **Reasoning Under Uncertainty** (Razonamiento en condiciones de incertidumbre), **Advanced Search** (Búsqueda avanzada), **Natural Language Processing** (Procesamiento del lenguaje natural), **Advanced Machine Learning** (Aprendizaje automático avanzado) y **Perception and Computer Vision** (Percepción y visión artificial). **Robotics**, la implementación física de algunos o todos los tópicos anteriores es referenciada en el sub campo de Robótica, destacada de manera central abajo en color verde oliva. (Aguirre, comunicación personal, 2024).

La Taxonomía presentada en la Figura 1 permite hacernos de una mirada del campo de IA, pero a la vez caracterizar el algoritmo ID3. Esto nos llevó a contestar la segunda pregunta de investigación.

P2: ¿En qué rama de IA se ubican los Árboles de Decisión?

A partir de los resultados obtenidos podemos decir que, los Árboles de Decisión y en particular el Algoritmo ID3 pertenecen al sub campo **Aprendizaje Automático Avanzado** (Advanced Machine Learning); y como tal requieren un **Aprendizaje Supervisado** (Supervised learning), en este sentido la base de conocimiento que sirve para entrenar el algoritmo posee los valores objetivo, lo que permite establecer a su vez una curva de aprendizaje que el científico debe observar y corregir de ser necesario.

FUNDAMENTOS LÓGICOS.

Entre sus distintos enfoques la Inteligencia Artificial tiene como objetivo la producción de sistemas que piensan racionalmente. El abordaje de este tipo de sistemas ha sido realizado implicando diferentes campos del conocimiento como la Lógica (Lógicas de primer orden, Lógicas Polivalentes, Sistemas axiomáticos, Deducción natural, Inducción, etc.), la Psicología (Constructivismos, Conductismo, etc.), la Biología (Neurología), Ingeniería del Conocimiento, Filosofía, Cibernética, Lingüística, etc. Si puntualizamos en la idea de realizar inferencias racionales, encontraremos que existen tres formas de razonamiento (Pierce, 1878):

- **Razonamiento deductivo**: Consistente en partir de la regla general, y que obtiene una conclusión a partir del caso particular -sustento de las ciencias formales-
- **Razonamiento inductivo**: que establece reglas generales, a partir de casos. –sustento de las ciencias fácticas-.
- **Razonamiento abductivos**: que permiten establecer hipótesis, a partir de casos particulares y reglas generales. –que es la forma auténtica de descubrimiento de nuevo conocimiento-

En todos ellos, los elementos que constituyen el razonamiento están dados por enunciados establecidos, podríamos decir perfectamente delimitados. Normalmente los agentes en el mundo real, no están provistos de elementos tan bien definidos; pero esto no es una limitación para realizar inferencias y que estas gocen de racionalidad.

La complejidad del mundo dificulta establecer consideraciones desprovistas de errores o aspectos faltantes; esta complejidad dificulta también establecer reglas generales para ser aplicadas, e incluso puede inhabilitar hipótesis propuestas.

Claro está, que la única excepción provenga de las ciencias formales, que pueden prescindir de una interpretación o correspondencia con el mundo real; incluso los dilemas y paradojas están restringidos en estos ámbitos. Podemos decir que la complejidad del mundo aporta incertidumbre, limitando la posibilidad de realizar inferencias racionales.

Es aquí donde la IA puede brindar soluciones computacionales a partir de métodos Inferenciales automatizados como los llamados Árboles de decisión, que pueden ser implementados como un componente fundamental en los sistemas que piensan racionalmente.

> La tarea, también llamada **aprendizaje inductivo**, consiste en aprender una función a partir de ejemplos de sus entradas y salidas…El aprendizaje inductivo consiste en encontrar una hipótesis consistente que verifique los ejemplos. La navaja de Ockham[1] sugiere elegir la hipótesis consistente más sencilla. (Russell y Norvig, 2004, p. 766)

Existen diversos criterios o algoritmos para implementar Árboles de decisión, uno de ellos vincula las nociones de Probabilidad y Entropía en el marco de la Teoría de la Información; este es caso de Iterative Dichotomiser 3 (ID3).

FUNDAMENTOS FÍSICO-MATEMÁTICO.

En esta sección vamos a establecer las ideas centrales que gobiernan la heurística del algoritmo ID3. Rudolf Julius Emanuel Clausius fue un Físico Alemán del siglo XIX, reconocido por ser uno de los "fundadores" de la Termodinámica (Meitner, 2017). En 1865 estableció la noción de **Entropía**(S) y una definición matemática de la misma como se observa en la Ecuación 1.

$$S = \frac{dQ}{T} \qquad (1)$$

[1] A William de Ockham (1280-1349), el filósofo más influyente de su siglo y un gran contribuidor a la epistemología medieval, la lógica y la metafísica, se le atribuye la afirmación denominada «La navaja de Ockham». Se expresa en latín "Entia non sunt multiplicanda praeter necessitatem", y en castellano «Las entidades no han de ser multiplicadas más allá de la necesidad». Desgraciadamente, este loable consejo no se encuentra en ninguna parte de sus escrituras con estas palabras precisamente (Russell y Norvig, 2004, p. 767).

Básicamente una razón entre la cantidad de calor (Q) y la temperatura absoluta (T). Famosa es su frase “La entropía del universo tiende a un máximo.” Para entender mejor esta noción podemos recurrir a una visión macroscópica estableciendo que “La Entropía es una medida cuantitativa del desorden...” (Sears, Zemansky, Young & Freedman, 2004).

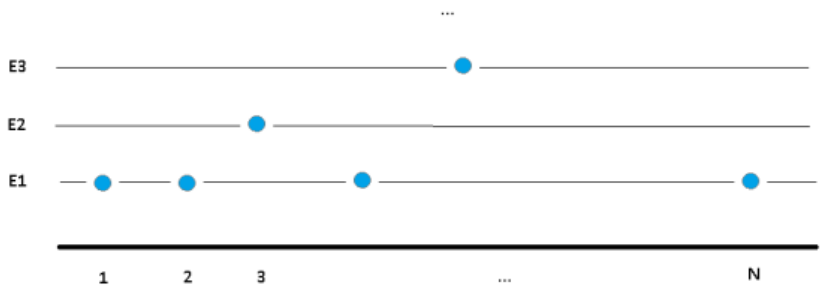
Podemos pensar en un gas más **desordenado** después de la expansión porque las moléculas se mueven en un volumen mayor y tienen más aleatoriedad de posición, para una referencia mayor ver **Apéndice I**. Más tarde y con base a las observaciones dadas por Maxwell, el Físico-Matemático Ludwig Edward Boltzmann demuestra en 1872 la siguiente identidad (Meitner, 2017) (Ecuación 2)

$$S = k \log w \qquad (2)$$

Mediante argumentaciones combinatorias y ciertos supuestos sobre la interacción molecular de un gas, establece que la entropía(S) se encuentra en función de los posibles estados microscópicos para un estado macroscópico dado (w), siendo K la constante de Boltzmann. Las ideas de Boltzmann fundan lo que se conoce como Física Estadística (Alonso y Finn, 1986). De manera resumida podemos pensar en un conjunto de N partículas que se encuentran en un gas ideal, como muestra la Figura 2.

Figura 2

Modelo molecular general de un gas ideal



Y establecer los siguientes supuestos

1. El número de partículas permanece constante durante todos los procesos que ocurren en el sistema.
2. Las partículas no interactúan, o solo lo hacen ligeramente.
3. El sistema está aislado, la energía total U debe ser constante; sin embargo, pueden cambiar la distribución de las partículas entre los estados disponibles de energía.

Dadas estas condiciones podemos reconocer que algunas partículas se encuentran en un estado energético, y otras en otro. Esto determina una partición del sistema de partículas que conforman el gas. Por consiguiente y mediante consideraciones fundamentalmente combinatorias, es posible establecer que tan probable es una partición dada mediante una distribución de Maxwell-Boltzmann dada en la Ecuación 3

$$P_K = \prod_{i=1}^{K} \frac{{g_i}^{n_i}}{n_i!} \tag{3}$$

Particularmente, cuando un sistema alcanza la partición más probable, se encuentra en un estado de **equilibrio estadístico**. Y es aquí donde el sistema posee máxima entropía. Un resultado notable determina que (Ecuaciones 4 y 5)

$$dA = \frac{dQ}{\mathrm{T}} \tag{4}$$

con

$$A = K \ln(P) = K\,N + K \ln\left(\frac{\mathrm{Z}}{\mathrm{N}}\right) N + \frac{U}{\mathrm{T}} \tag{5}$$

La identidad presentada en las Ecuaciones 4 y 5 puede ser ampliada en el Apéndice II. La importancia de esta identidad en nuestra exposición radica en proveernos de un instrumento matemático para medir la

entropía en un sistema y en particular establecer cuando estamos en un sistema con equilibrio estadístico o con máxima entropía.

En 1948 el genial matemático e ingeniero Claude Elwood Shannon redactó "**A Mathematical Theory of Communication**" en "The Bell System Technical Journal". La importancia del texto es conocida por todos ya que en él se funda la llamada Era de la Información. Una de las preocupaciones centrales de Shannon, expresada en este documento, era establecer la confiabilidad de los mensajes durante los procesos de comunicación: "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" (Shannon, 1948, p. 379). Nos interesa por tanto puntualizar en el siguiente resultado que hemos decidido mostrar de manera literal en la Figura 3

Figura 3

Formulación original presenta en la página 19, del famoso escrito de Shannon del año 1948

Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant.

En el de manera brillante se conectan las ideas de entropía e información, al considerar los elementos de información supeditados a condiciones semejantes a las partículas en los gases. De hecho, demuestra que el operador H es análogo al conocido resultado dado por Boltzmann en su Mecánica Estadística para la entropía. Es de notar que p_i es la probabilidad de que una cierta unidad de información se encuentre en la celda i (léase, adquiera un valor específico). Y para nosotros es sumamente importante a la vez que conveniente limitarnos a la Ecuación 6, considerando K como la unidad

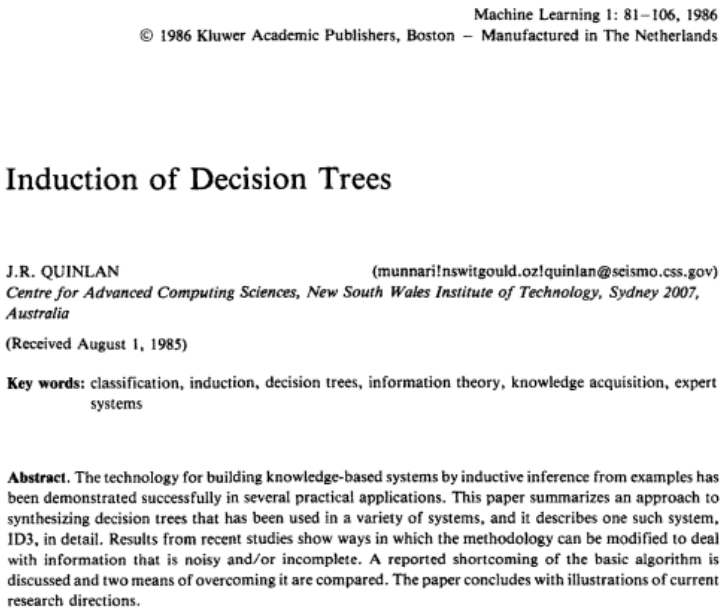
$$\mathrm{H(x)} = -\mathrm{p(x)}\log\big(\mathrm{p(x)}\big) - \mathrm{q(x)}\log\big(\mathrm{q(x)}\big) \qquad (6)$$

La Ecuación (6) para Shannon es una medida de información, elección e **incertidumbre**, este último concepto nos será especialmente útil, y puede ser ampliado en el Apéndice III.

Finalmente presentamos el último eslabón en nuestra red de conexiones. En el año 1979 el Ingeniero John Ross Quinlan presenta "INDUCTION OVER LARGE DATABASES" en la Universidad de Stanford. En este trabajo alude a un procedimiento basado en inducción a partir del cual es posible extraer reglas de aprendizaje representadas en un Árbol de decisión (Quinlan, 1979). No obstante, vamos a remitirnos al informe presentado en el año 1986 conocido como "Induction of Decision Trees", cuya portada puede verse en la Figura 4.

Figura 4

Portada del escrito "Induction of Decision Trees", presentado por Quinlan en el año 1986



Machine Learning 1: 81–106, 1986

Induction of Decision Trees

J.R. QUINLAN (munnari!nswitgould.oz!quinlan@seismo.css.gov)
Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007, Australia

(Received August 1, 1985)

Key words: classification, induction, decision trees, information theory, knowledge acquisition, expert systems

Abstract. The technology for building knowledge-based systems by inductive inference from examples has been demonstrated successfully in several practical applications. This paper summarizes an approach to synthesizing decision trees that has been used in a variety of systems, and it describes one such system, ID3, in detail. Results from recent studies show ways in which the methodology can be modified to deal with information that is noisy and/or incomplete. A reported shortcoming of the basic algorithm is discussed and two means of overcoming it are compared. The paper concludes with illustrations of current research directions.

La razón de ello es la clara intención de divulgar su algoritmo ID3, que permite la construcción de árboles de decisión por medio de la inducción: “The induction task is to develop a classification rule that can determine the class of any object from its values of the attributes” (Quinlan, 1986, p. 86).

Es posible considerar la Ecuación 6 para calcular la entropía (I) de un elemento de información llamado “atributo”, a partir de una clase que posee dos valores (digamos positivos y negativos); este atributo está constituido a su vez de elementos que la caracterizan, y que permiten vía la Ecuación 6 calcular una suerte de entropía esperada (E). Surge de esta consideración la posibilidad de definir una heurística para la construcción de reglas, conocida como Ganancia de Información(G) (Ecuación 7)

$$\mathrm{G(a)} = \mathrm{I(a)} - \mathrm{E(a)} \qquad (7)$$

A partir de la Ecuación 7 podremos calcular la ganancia de información de un atributo, al comparar la incertidumbre obtenida a partir de su elección, contra la incertidumbre esperada de sus elementos. Y construir, por tanto, con este procedimiento, un árbol de decisión comenzando por la raíz.

DESCRIPCIÓN GENERAL DEL ALGORÍTMO ID3

Hemos podido en términos generales establecer las bases teóricas que hacen que el algoritmo ID3 funcione. Trataremos ahora de manera resumida de presentar el algoritmo, en primer lugar, aludiendo al formato de su base de conocimiento y más tarde al algoritmo en sí.

Base de conocimiento, conjunto de instancias.

Quinlan (1986) en la primera parte de su escrito requiere de un ejemplo tutorial preparado por un experto del dominio: “…set of tutorial examples prepared by a domain expert…” (Quinlan, 1986, p. 84) para entrenar el algoritmo. En particular propone a modo de ejemplo un dataset, que es favorito en la presentación del algoritmo ID3, como puede ser visto en la

Figura 5

Data Set de entrenamiento "weather", propuesto en el escrito "Induction of Decision Trees", presentado por Quinlan en el año 1986

Table 1. A small training set

No.	Attributes				Class
	Outlook	Temperature	Humidity	Windy	
1	sunny	hot	high	false	N
2	sunny	hot	high	true	N
3	overcast	hot	high	false	P
4	rain	mild	high	false	P
5	rain	cool	normal	false	P
6	rain	cool	normal	true	N
7	overcast	cool	normal	true	P
8	sunny	mild	high	false	N
9	sunny	cool	normal	false	P
10	rain	mild	normal	false	P
11	sunny	mild	normal	true	P
12	overcast	mild	high	true	P
13	overcast	hot	normal	false	P
14	rain	mild	high	true	N

Tal base de conocimiento está compuesta de "**Instancias**" (catorce en nuestro ejemplo); se reconocen también otras categorías conocidas como "**Atributos**" (a saber, Outlook, Temperature, Humidity, Windy), descriptos en términos de "**Elementos**" (en el caso Outlook, tendremos sunny, overcast y rain).

En las instancias de la aplicación existe una categoría llamada "**Class**", compuesta solamente de dos elementos, uno **positivo(P)** y otro **negativo(N)**. Esta categoría es central en la heurística del algoritmo, pues a partir del uso de "ventanas" es posible decidir entre distintos atributos evaluando la probabilidad de positivo y negativo. Esto permite determinar la ganancia de información del atributo.

Quinlan advierte sobre no considerar instancias idénticas que tengan elementos de clase opuestos: "…if the training set contains two objects that have identical values for each attribute and yet belong to different classes, it is clearly impossible to differentiate between these objects with reference only to the given attributes" (Quinlan, 1986, p. 86).

El último punto que señalaremos consiste en la **iteración** sobre un sub conjunto del conjunto de entrenamiento dado, que podrá ser ampliado conforme el árbol (podríamos decir la teoría) realice mejores predicciones.

Algoritmo ID3.

La implementación del algoritmo ID3, consiste en aplicar una técnica **recursiva** de programación: "ID3 examines all candidate attributes and chooses A to maximize gain(A), forms the tree as above, and then uses the same process recursively to form decision trees for the residual subsets C_1, C_2 . . .. C_v." (Quinlan, 1986, p. 90)

De esta manera en nuestro ejemplo puede ser elegido un atributo, digamos "Outlook". Para el que podemos calcular su entropía (o incertidumbre) en los términos dados por la Ecuación 6 al tener en cuenta que p es la probabilidad de positivos, y q la de negativos (Ecuación 8).

$$I_{\text{Outlook}}(9,5) = \left(-\frac{9}{14}\right)\log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right)\log_2\left(\frac{5}{14}\right) = 0.9402 \quad (8)$$

Por otro lado, es posible calcular la **entropía esperada del atributo**, a partir de la entropía de cada uno de sus elementos. En nuestro caso resultan en (Ecuación 9)

$$I_{sunny}(2,3) = \left(-\frac{2}{5}\right)\log_2\left(\frac{2}{5}\right) - \left(\frac{3}{5}\right)\log_2\left(\frac{3}{5}\right) = 0.9709 \quad (9)$$

Aquí sunny, como muestra la Ecuación 9, posee máxima entropía y por consiguiente la mayor incertidumbre. Overcast, presenta entropía nula como muestra la Ecuación 10 y por consiguiente nula incertidumbre; ciertamente en este caso podemos estar seguros que podemos concluir en el elemento de clase "positivo" (Ecuación 11).

$$I_{overcast}(4,0) = \left(-\frac{4}{4}\right)\log_2\left(\frac{4}{4}\right) - \left(\frac{0}{4}\right)\log_2\left(\frac{0}{4}\right) = (-1)\ 0 - [0][-\infty] = 0 \quad (10)$$

$$I_{rain}(3,2) = \left(-\frac{3}{5}\right)\log_2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right)\log_2\left(\frac{2}{5}\right) = 0.9709 \qquad (11)$$

Finalmente, rain, presenta máxima entropía, y por consiguiente máxima incertidumbre. Estos valores permiten calcular la **entropía esperada** del atributo Outlook (Ecuación 12):

$$\mathrm{E}(\mathrm{Outlook}) = \frac{5\ \ 0.9709\ +\ 4\ \ 0\ +5\ \ 0.9709}{14} = 0.6935 \qquad (12)$$

Para explicarlo en términos simples, la entropía de la condición Outlook, implica una incertidumbre de 0.6935 a partir de considerar sus elementos. Finalmente, estos valores permiten calcular la **Ganancia de información** del atributo Outlook (Ecuación 13):

$$G(\mathrm{Outlook}) = I_{\mathrm{Outlook}}(9{,}5) - \mathrm{E}(\mathrm{Outlook}) = 0.9402 - \ 0.6935 = \mathbf{0.2467} \qquad (13)$$

Básicamente la Ganancia de información permite considerar cuanta incertidumbre estamos dispuestos a aceptar al optar por el atributo Outlook como un nodo de nuestro árbol, con base a las opciones (o elementos) a los que estaremos supeditados. Ciertamente la máxima ganancia de información respecto de los otros atributos se da para Outlook.

Este procedimiento puede ser reproducido de manera recursiva, teniendo en cuenta como condición de terminación establecer los elementos de clase como se sugirió en la ecuación (10). El árbol resultante permite construir una “teoría” suficientemente buena, con un costo computacional bajo, al realizar una inducción sobre las instancias del elemento.

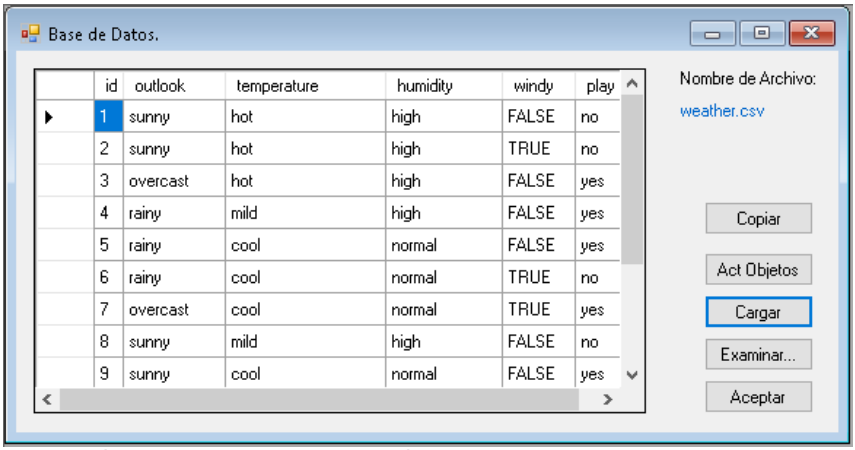
> The worth of ID3's attribute-selecting heuristic can be assessed by the simplicity of the resulting decision trees, or, more to the point, by how well those trees express real relationships between class and attributes as demonstrated by the accuracy with which they classify objects other than those in the training set (their predictive accuracy) (Quinlan, 1986, p. 91)

IMPLEMENTACIÓN COMPUTACIONAL

Finalmente se muestra una implementación del algoritmo en un lenguaje C#, empleando memoria dinámica para la construcción del árbol (Deitel & Deitel, 2008).

Figura 6

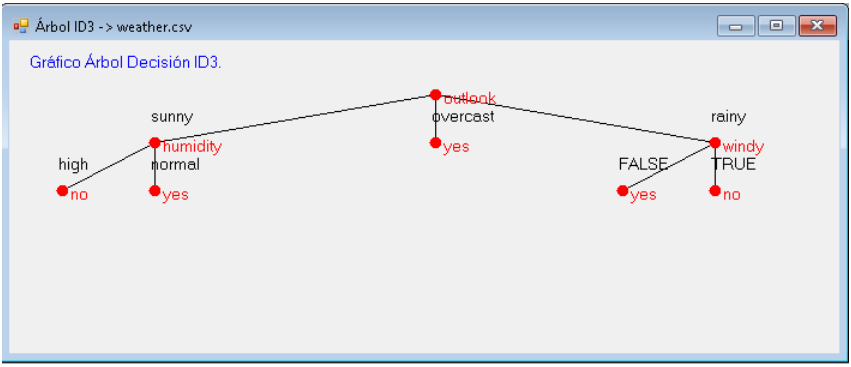
Acceso al Data Set de entrenamiento "weather.csv", con las catorce instancias propuestas por Quinlan



Nota. El gráfico corresponde a una aplicación propia implementada en C#, con la habilidad de leer archivos de tipo .csv. Y poner disponible las instancias, para la construcción en memoria dinámica del árbol ID3.

Figura 7

Árbol de decisión ID3, obtenido a partir de la base de conocimiento "weather.csv"

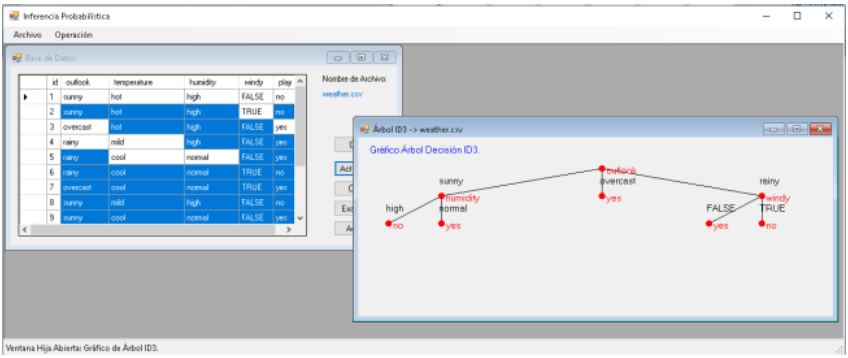


Nota. El gráfico corresponde a una aplicación propia implementada en C#, en el pueden verse los atributos Humidity, Outlook, Windy y sus respectivos elementos. El resultado es idéntico al obtenido por Quinlan con su famoso data set.

Como estrategia de construcción del árbol de decisión, se implementó la ejecución del algoritmo ID3 a la vez que se calculaban los puntos necesarios para su representación gráfica. Y siempre apelando a la técnica recursiva. (Aguirre, comunicación personal, 2024)

Figura 8

Interfaz completa donde se observa el árbol de inducción logrado a partir del algoritmo ID3, junto con la base de conocimiento "weather.csv"-luego de reconocer los elementos de cada atributo- empleando child windows.



CONCLUSIONES

Existe una notable interconexión entre los distintos campos del conocimiento (Lógica, Matemática, Filosofía, Física, Historia, Biología, etc.) al postular un modelo de la realidad. Este trabajo de manera resumida ha permitido experimentar este proceso, un privilegio y resultado de meses de ensayo y error, replanteo, investigación documental, horas de programación, etc. El estudio de modelos de la realidad, que en particular tengan implementaciones en modelos computacionales es un objetivo logrado…seguirán otros en el futuro.

AGRADECIMIENTOS

Los autores de este trabajo desean agradecer a la organización de ECEFI 2024 por generar un espacio para la presentación de trabajos de investigación. Al Doctor Carlos Neil, por incentivarme y acompañarme en el nacimiento de un investigador novel. A mi familia por tolerar ausencias mentales…y a veces físicas.

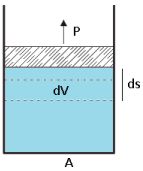
APÉNDICE I

Notas del libro de Sears, Zemansky, Young y Freedman (pág. 754)

Entropía: “es una medida cuantitativa del desorden.”

Consideremos una expresión isotérmica infinitesimal de un gas ideal (y dada la primera ley de la termodinámica “…el calor(Q) neto que fluye hacia la máquina en un proceso cíclico es igual al trabajo(W) realizado por la máquina” $U_1 – U_2 = 0 = Q – W \rightarrow Q = W$)

$$dU = 0 \rightarrow dQ = dW = d(f \times s) = d(p \times A \times s) =$$
$$= p\ x.\ A\ x\ ds = p\ .dV = \frac{n\,R\,T}{V}\ dV$$



T, es constante debido al proceso isotérmico.

U_G, es constante debido a que la energía interna del gas depende de su temperatura, con lo cual dU = 0.

Debido al tratamiento infinitesimal, la presión p puede ser considerada constante.
Recordemos la ecuación fundamental del gas ideal $p\,V = n\,R\,T$, donde p: presión, n: número de moles, R: constante molar de los gases, T: temperatura.
De manera que

$$\frac{\boldsymbol{dQ}}{n\,R\,\boldsymbol{T}} = \frac{dV}{V}$$

"El gas está en un estado más **desordenado** después de la expansión porque las moléculas se mueven en un volumen mayor y tienen más aleatoriedad de posición."

"…los procesos se efectúan naturalmente en la dirección de desorden creciente,

(1) La adición de calor a un cuerpo aumenta su desorden (aumenta su vibración molecular)
(2) La expansión libre de un gas (aumenta la distancia entre moléculas)"

Matemáticamente, **definimos la entropía(S)** en un proceso reversible infinitesimal a temperatura T como

$$dS = \frac{dQ}{T} \qquad (dS = n\,R\,\frac{dV}{V})$$

Para una cantidad discreta de calor Q (intercambio de energía entre el sistema y el entorno)

$$\Delta S = S_2 - S_1 = \frac{Q}{T} \qquad \text{[S]: J/K}$$

Apéndice II

Enfoque molecular basado en un ejercicio propuesto en el Libro de Física de Alonso y Finn (Ejercicio 11.3, pág. 490).

En un sistema en equilibrio estadístico y que obedece a la estadística de Maxwell-Boltzmann.

$$A = K\ln(P) = K\left(N - \sum_i n_i \ln\left(\frac{n_i}{g_i}\right)\right) = K\,N - K\sum_i n_i \ln\left(\frac{n_i}{g_i}\right)$$

$$= K\,N - K\sum_i n_i\ \ln\left(\frac{\mathrm{g_i}}{\mathrm{Z}}\frac{\mathrm{N}}{g_i}\mathrm{e}^{-\beta \mathrm{E_i}}\right) =$$

$$K\,N - K\sum_i n_i\ \ln\left(\frac{\mathrm{N}}{\mathrm{Z}}\mathrm{e}^{\left(-\frac{\mathrm{E_i}}{\mathrm{K\,T}}\right)}\right) = K\,N - K\sum_i n_i\ \left(\ln\left(\frac{\mathrm{N}}{\mathrm{Z}}\right) - \frac{\mathrm{E_i}}{\mathrm{K\,T}}\right) =$$

$$= K\,N - K\ln\left(\frac{\mathrm{N}}{\mathrm{Z}}\right)\sum_i n_i + K\frac{1}{\mathrm{K\,T}}\sum_i \mathrm{E_i} = K\,N - K\ln\left(\frac{\mathrm{N}}{\mathrm{Z}}\right)N + \frac{1}{\mathrm{T}}U$$

$$= K\,N + K\ln\left(\frac{\mathrm{Z}}{\mathrm{N}}\right)N + \frac{U}{\mathrm{T}}$$

De manera que

$$\boldsymbol{A = K\ln(P) = K\,N + K\ln\left(\frac{\mathrm{Z}}{\mathrm{N}}\right)N + \frac{U}{\mathrm{T}}} \qquad \text{(Ec. 11.28)}$$

Para una transformación reversible infinitesimal en la cual el número total de partículas no varía.

$$\frac{dA}{dT} = \frac{d\left(K\,N + K\,N\ln\left(\frac{\mathrm{Z}}{\mathrm{N}}\right) + \frac{U}{\mathrm{T}}\right)}{dT}$$

$$\frac{dA}{dT} = 0 + K\,N\frac{1}{N}\frac{\mathrm{N}}{\mathrm{Z}}\frac{\mathrm{dZ}}{\mathrm{dT}} + \frac{dU}{\mathrm{dT}}\frac{1}{\mathrm{T}} - \frac{\mathrm{U}}{\mathrm{T^2}}$$

$$\boldsymbol{dA = K\,N\frac{\mathrm{dZ}}{\mathrm{Z}} + \frac{dU}{\mathrm{T}} - \frac{\mathrm{U}}{\mathrm{T^2}}dT} \qquad \text{(Ec. 11.38)}$$

Sea la función de partición, dada en términos de la temperatura absoluta.

$$Z = \sum_{i=1}^{K} g_i e^{-\frac{E_i}{KT}}$$

Diferenciando

$$dZ = d\left(\sum_{i} g_i e^{-\frac{E_i}{KT}}\right)$$

$$dZ = \sum_{i} g_i \, d\left(-\frac{E_i}{KT}\right) e^{-\frac{E_i}{KT}} = \sum_{i} g_i \left(-\frac{dE_i}{KT} + \frac{E_i}{KT^2} dT\right) e^{-\frac{E_i}{KT}} =$$

$$= \sum_{i} g_i \left(-\frac{dE_i}{KT}\right) e^{-\frac{E_i}{KT}} + \sum_{i} g_i \left(\frac{E_i}{KT^2}\right) e^{-\frac{E_i}{KT}} dT$$

Tras lo cual

$$dZ = \sum_{i} g_i \left(-\frac{dE_i}{KT}\right) e^{-\frac{E_i}{KT}} + \sum_{i} g_i \left(\frac{E_i}{KT^2}\right) e^{-\frac{E_i}{KT}} dT$$

De esta última ecuación al multiplicar por un factor conveniente, resulta que

$$K\,N\frac{dZ}{Z} = \sum_{i} g_i \frac{K\,N}{Z}\left(-\frac{dE_i}{KT}\right) e^{-\frac{E_i}{KT}} + \sum_{i} g_i \frac{K\,N}{Z}\left(\frac{E_i}{KT^2}\right) e^{-\frac{E_i}{KT}} dT$$

$$K\,N\frac{dZ}{Z} = \left(-\frac{1}{T}\right)\sum_{i} g_i \frac{N}{Z} e^{-\frac{E_i}{KT}} dE_i + \frac{1}{T^2}\sum_{i} g_i \frac{N}{Z} E_i e^{-\frac{E_i}{KT}} dT$$

$$K\,N\frac{dZ}{Z} = \left(-\frac{1}{T}\right)\sum_{i} n_i \, dE_i + \frac{1}{T^2}\sum_{i} n_i \, E_i \, dT$$

Por lo tanto

$$K\,N\frac{dZ}{Z} = \frac{1}{T}\, dW + \frac{1}{T^2} U \, dT$$

Retomando de la ecuación (Ec. 11.38)

$$dA = K\,N\frac{\mathrm{dZ}}{\mathrm{Z}} + \frac{dU}{\mathrm{T}} - \frac{\mathrm{U}}{\mathrm{T}^2}dT = \left(\frac{1}{\mathrm{T}}\,dW + \frac{1}{\mathrm{T}^2}\mathrm{U}\,\mathrm{dT}\right) + \frac{dU}{\mathrm{T}} - \frac{\mathrm{U}}{\mathrm{T}^2}dT =$$
$$= \frac{dW}{\mathrm{T}} + \frac{dU}{\mathrm{T}} = \frac{dW + dU}{\mathrm{T}} = \frac{dQ}{\mathrm{T}}$$

Por lo cual $\boldsymbol{dA = \frac{dQ}{\mathrm{T}}}$, que es la **definición de entropía con $\boldsymbol{A = K\ln(P)}$**

Apéndice III

● Interpretación de la entropía H.

La siguiente imagen se extrae directamente de la p. 20 del texto original publicado por Shannon en el año 1948. En ella se muestra la gráfica de la entropía en función de una probabilidad p de cierto sistema con q = (1 – p)

The entropy in the case of two possibilities with probabilities p and $q = 1 - p$, namely

$$H = -(p \log p + q \log q)$$

is plotted in Fig. 7 as a function of p.

The quantity H has a number of interesting properties which further substantiate it as a reasonable measure of choice or information.

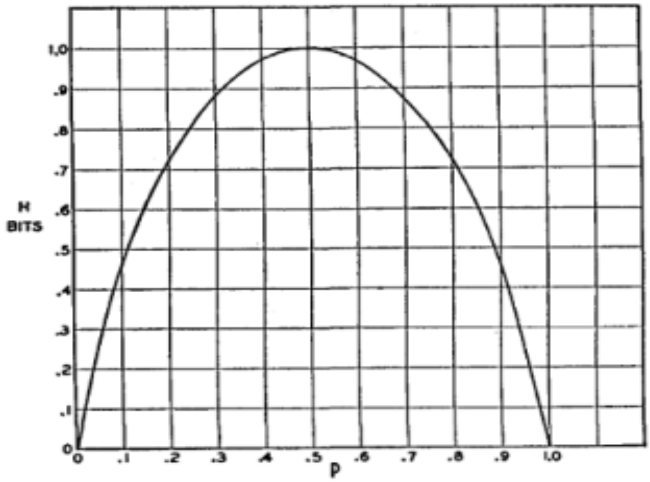
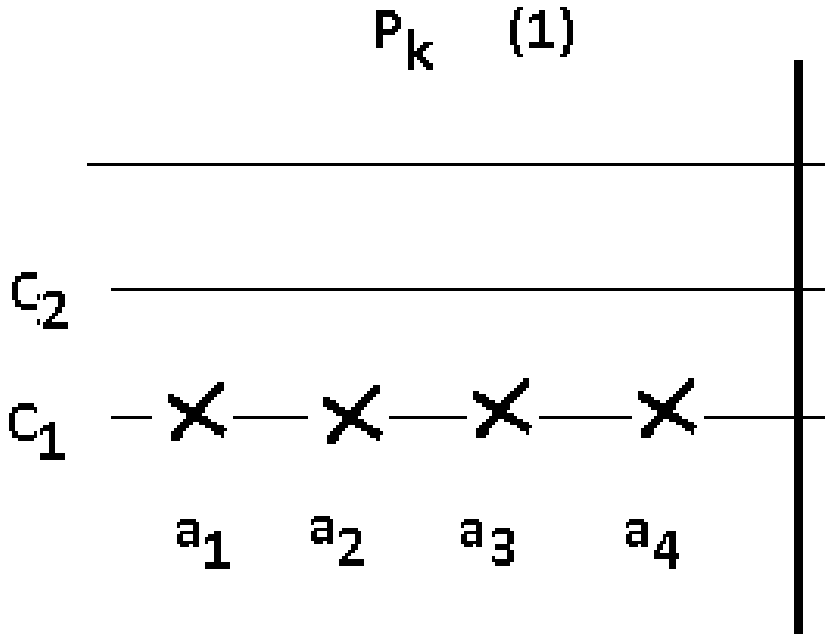


Fig. 7—Entropy in the case of two possibilities with probabilities p and $(1 - p)$.

Para explicar el funcionamiento de la entropía H y con vista a nuestros propios fines recurriremos a ejemplos particulares.

√ Pensemos en cuatro elementos de información a_1, a_2, a_3, a_4. Supondremos que ellos pueden tomar dos estados c_1 y c_2 (digamos "bajo" 0 y "alto" 1) como muestra el siguiente gráfico



Desde un punto de vista físico podríamos decir que el sistema presenta una determinada partición ("todos ceros"). En tal caso la probabilidad de "todos ceros" sería

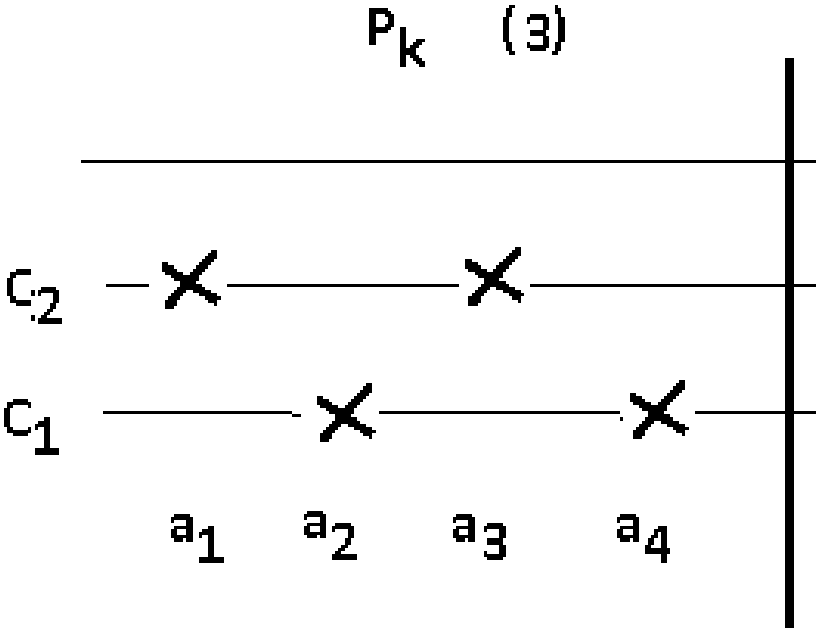
$$p = P[a_1 = 0 \wedge a_2 = 0 \wedge a_3 = 0 \wedge a_4 = 0] = 1 \quad y \quad q = (1 - p)$$

Con lo cual la evaluación de la expresión de entropía (con la adecuación matemática necesaria) sería

$$H(\text{"todos ceros"}) = -p(\text{"todos ceros"})\log\big(p(\text{"todos ceros"})\big) - q(\text{"todos ceros"})\log\big(q(\text{"todos ceros"})\big) = 0$$

Es decir, la entropía será nula, hay plena certeza de que los elementos de información son todos ceros –no hay incertidumbre en los términos de Shannon-.

√ Consideremos ahora que estos cuatro elementos de información a_1, a_2, a_3, a_4 adquieren una determinada configuración o partición. Llamemos a la misma "dos ceros"



Podríamos definir en tal caso la probabilidad de "dos ceros" (la probabilidad de esta partición) como

$$p = P[a_1 = 1 \wedge a_2 = 0 \wedge a_3 = 1 \wedge a_4 = 0] = \frac{1}{2} \quad y \quad q = (1 - p) = \frac{1}{2}$$

Con lo cual la evaluación de la expresión sería

$$\mathrm{H}(\text{"dos ceros"}) = -\mathrm{p}(\text{"dos ceros"})\log\big(\mathrm{p}(\text{"dos ceros"})\big) - \mathrm{q}(\text{"dos ceros"})\log\big(\mathrm{q}(\text{"dos ceros"})\big) = \frac{1}{2}$$

Es decir, la entropía sería 0.5, nunca podríamos estar seguros de sí un elemento de información alcanza el estado cero -en tal configuración, el sistema se encuentra en equilibrio estadístico, puesto que la entropía es máxima-.

Apéndice IV

● **Una aplicación alternativa del algoritmo ID3.**

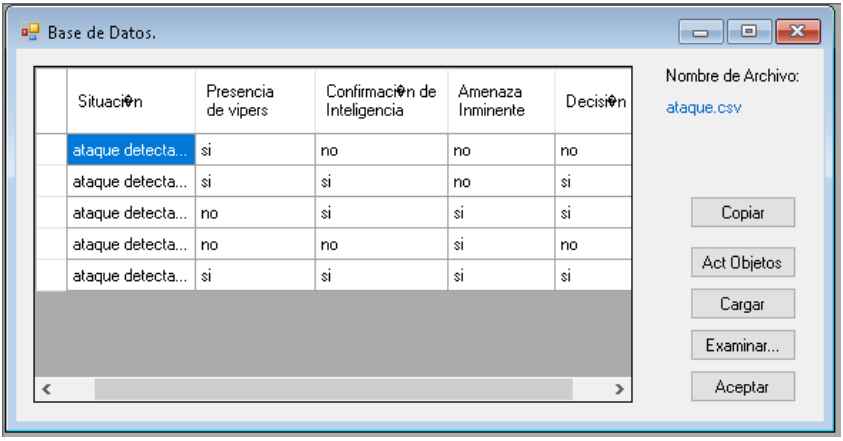
Informe de aplicaciones en Inteligencia: Toma de decisiones a nivel operacional.

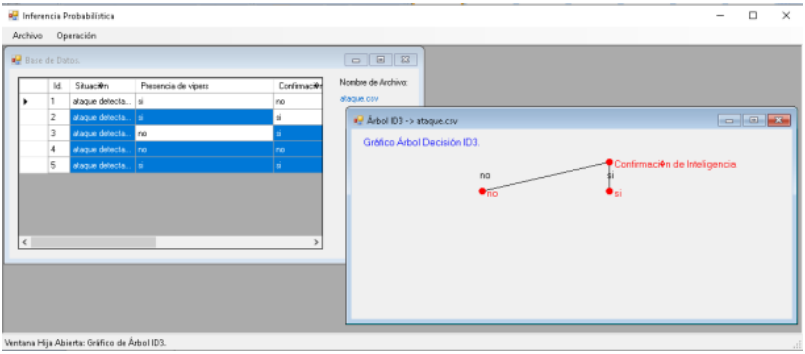
Tabla de datos: instancias de la base de conocimiento

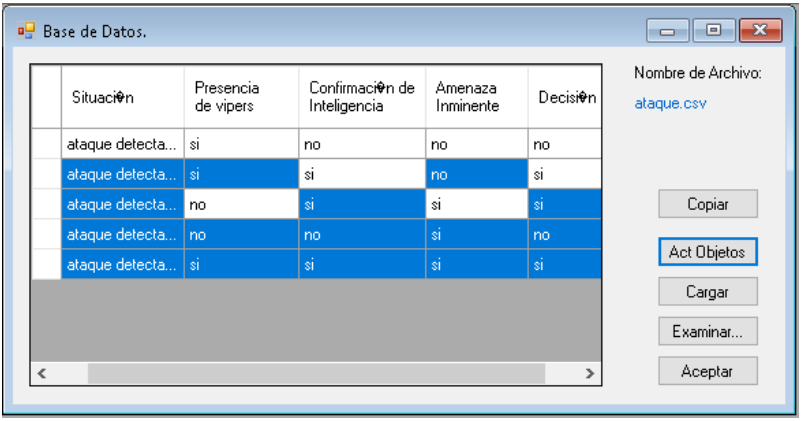
Id.	Situación	Presencia de vipers	Confirmación de Inteligencia	Amenaza Inminente	Decisión
1	ataque detectado	si	no	no	no
2	ataque detectado	si	si	no	si
3	ataque detectado	no	si	si	atacar
4	ataque detectado	no	no	si	no
5	ataque detectado	si	si	si	atacar

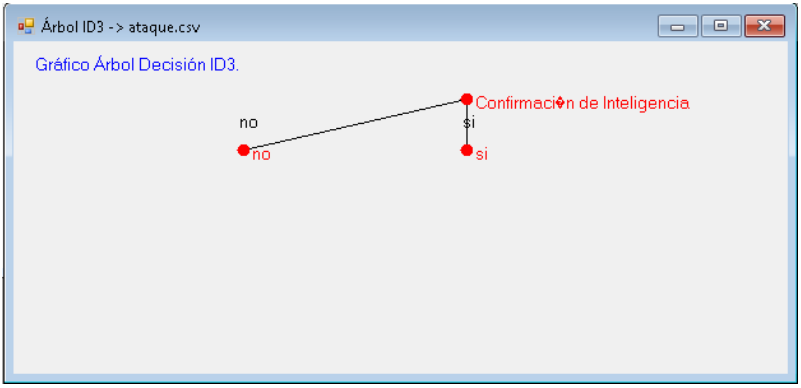
Ingeniería de Conocimiento: Se solicita a Bin que cree una base de conocimiento para aplicar algoritmo ID3.
Ejecución de Matía Gael Aguirre-Polenta.

Implementación: con C#, vista de resultados obtenidos




Inferencia Probabilística
Archivo
Operación
Base de Datos.
Id.
Situación
Presencia de vipers
Confirmación
Nombre de Archivo:
ataque.csv
Árbol ID3 -> ataque.csv
Gráfico Árbol Decisión ID3.
Confirmación de Inteligencia
no
si
Ventana Hija Abierta: Gráfico de Árbol ID3.


Base de Datos.
Situación
Presencia de vipers
Confirmación de Inteligencia
Amenaza Inminente
Decisión
ataque detecta... si no no no
ataque detecta... si si no si
ataque detecta... no si si si
ataque detecta... no no si no
ataque detecta... si si si si
Nombre de Archivo:
ataque.csv
Copiar
Act Objetos
Cargar
Examinar...
Aceptar


Árbol ID3 -> ataque.csv
Gráfico Árbol Decisión ID3.
Confirmación de Inteligencia
no
si
no
si

REFERENCIAS

Alonso, M., & Finn, Edward J. (1986). *Física. Fundamentos Cuánticos y Estadísticos* (C. A. Heras y J. A. Barreto Araujo, Trans.). (Vol. 3). Addison-Wesley Iberoamericana S. A. (Original work published 1968).

Audesirk, T., Audesirk, G., & Byers, B. E. (2008). *Biología: La vida en la tierra* (A. V. Flores Flores, Trans.). (8va ed.). Pearson Educación de México.

Deitel, H. M., & Deitel, P. J. (2007). *Cómo programar en C#* (2da ed.). Pearson Education de México S.A.

ACM & IEEE (2013). *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. ACM. https://doi.org/10.1145/2534860

Meitner, L. (2017). *Teorema H*. https://www.fisica.uns.edu.ar

Quinlan, J. R. (1979). *Induction over Large Databases*. Stanford University.

Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, *1*, 81–106. https://doi.org/10.1007/BF00116251.

Real Academia Española. (s.f.). Inteligencia. En Diccionario de la lengua española. Recuperado el 15 de enero, 2024, de https://dle.rae.es/inteligencia.

Russell, B. (1983). *El conocimiento humano* (N. Miguez, Trans.). Ediciones Orbis S.A. (Original work published 1948).

Russell, S. J. & Norvig, P. (2004). *Inteligencia artificial. Un enfoque moderno* (J. M. l. Corchado Rodríguez, F. M. Rubio, J. M. Cadenas Figueredo, L. D. Hernández Molinero, E. P. Arís, R. Fuentetaja Pinzán, M. Robledo de los Santos & R. Rizo Aldeguer, Trans.). (2nd ed.). Pearson Educación de Madrid.

Sears, F. W., Zemansky. M. W., Young, H. D., & Freedman, R. A. (2004). *Física Universitaria* (A. E. Brito, Trans.). (Vol. 1). (10th ed.). Pearson.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Labs Technical Journal*, *27*(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Peirce, C. S. (1994). *Collected papers of Charles Sanders Peirce* (Charles Hartshorne, Paul Weiss & Arthur W. Burks, Ed.). InteLex Corporation. (Original work published 1931-1935 (Vol. I - VI); 1958 (Vol. VII and VIII)). https://search.worldcat.org/title/collected-papers-of-charles-sanders-peirce/oclc/503075225

.

* * *