

Sinergias entre la Toma de Decisiones y la Calidad de las Informaciones

Alonso Secades, Vidal

Escuela Universitaria de Informática
Universidad Pontificia de Salamanca

López Rivero, Alfonso José

Escuela Universitaria de Informática
Universidad Pontificia de Salamanca

Joyanes Aguilar, Luis

Director del Departamento de Lenguajes, Sistemas Informáticos e Ingeniería del Software
Universidad Pontificia de Salamanca. Campus de Madrid

Abstract

Este artículo describe las relaciones e interacciones de las informaciones utilizadas en toma de decisiones mostrando como a distintos niveles de calidad de las informaciones las tomas de decisiones serán variables. Se analizarán mediante métodos cuantitativos las sinergias producidas entre tomas de decisiones y calidad de informaciones.

El análisis partirá de muestras aleatorias de datos con diferentes niveles de calidad. Además se tendrá en cuenta en el análisis multivariante los distintos equipos de toma de decisiones evitando los efectos colaterales que puedan producir factores de ruido en el análisis en estudio.

Para reflejar la calidad de los datos almacenados se precisa establecer algún formato adjunto, tal que el tratamiento posterior pueda interpretarse junto con la calidad del mismo. Así, la calidad de las informaciones quedará reflejada y se aplicará en la toma de decisiones. Las conclusiones demostrarán que existe sinergia entre la calidad de las informaciones y la toma de decisiones.

1. Introducción

Uno de los principales aspectos que deben afrontar los profesionales informáticos en la elaboración de sistemas de información empleados en la toma de decisiones estriba en el diseño de las bases de datos donde va a residir la información a utilizar en la toma de decisiones.

Este diseño se desarrolla de acuerdo a las técnicas y procedimientos estandarizados existentes, que permiten garantizar la integridad de los datos almacenados en la base de datos de tal forma que se pueda asegurar la calidad y exactitud de los mismos.

A pesar de la experiencia acumulada en aseguramiento de la calidad de datos, en la actualidad todavía continúan existiendo problemas. Esto es debido a que a la hora de realizar el diseño de una base de datos, los diseñadores de bases de datos deben contemplar las guías proporcionadas por la directiva para las tomas de decisiones, incluso más que las decisiones propias derivadas de las funciones a realizar por el sistema de información de la empresa.

La principal implicación de estas guías en el diseño se refleja en que los datos en estos sistemas no deben ser de calidad ideal ya que la calidad de las informaciones puede ser contemplada desde diferentes puntos de vista en función de las necesidades de cada departamento. Esto provoca que los diseñadores de bases de datos se encuentran con una dimensión relativa de la calidad que podrá ser diferente en los departamentos de la empresa.

Frecuentemente, los directivos deben tomar decisiones partiendo de las imperfecciones presentes en los datos de la base de datos. Para compensar estas imperfecciones, se debe efectuar la toma de decisiones estando al corriente de la idiosincrasia de los datos.

Un grave problema es que este conocimiento intuitivo se ha perdido, ya que los datos se utilizan, normalmente, por distintas partes y usuarios. Estos posibles usuarios que no poseen una destreza intuitiva para el tratamiento para los datos, están forzados a aceptar los datos tal y como le son presentados. Esta aceptación implica asumir que todos los valores de datos son válidos por igual, para evitar utilizar datos cuya calidad no pueda ser garantizada directamente por ellos. Esto conlleva a que algunos sistemas de soporte de decisiones no se pueden utilizar completamente.

El sentido común afirma que el conocimiento atendiendo a la calidad de datos, debe ser utilizado para lograr un beneficio, siendo esta proposición la que se plantea examinar. Además se debe tener en cuenta la clase de información a considerar sobre la calidad de datos, la cual es más apropiado que dependa de la naturaleza del proceso de decisión que de los propios datos. Por otra parte, la eficacia de la información que define la calidad de datos depende en gran medida de la satisfacción del usuario.

El objetivo es ayudar a los diseñadores de bases de datos encaminadas a la toma de decisiones, a proporcionarles una estructura y unos conceptos consistentes. Partiendo de la estructura y conceptos proporcionados se podrá establecer un punto de partida que determine el tipo y formato de información necesaria, atendiendo a la calidad de los datos, que sea más efectiva en cada situación. Un diseñador de base de datos cuya función sea incluir información según la calidad de datos necesita saber que beneficios se obtienen del sistema de información.

Este artículo pretende afrontar este problema para analizar el impacto de la información asociada con la calidad de los datos, presentándolo con datos actuales y de acuerdo a la calidad de datos.

En particular, se procederá a la elección del alquiler de un piso entre varias alternativas teniendo en cuenta la presencia de informaciones acerca de la calidad de los datos. Estas informaciones vendrán reflejadas en diferentes formas y contextos permitiendo evaluar la toma de esta decisión desde datos de baja calidad hasta datos con alta calidad. Los datos necesarios para realizar el experimento han sido facilitados por grupos de trabajo de la Escuela Universitaria de Informática.

2. Estado del arte

No es posible afirmar que el estudio sistemático de la calidad de los datos y de la información sea un fenómeno que haya sido estudiado a lo largo de mucho tiempo, ya que se trata de un fenómeno relativamente reciente.

Entre los estudios existentes, destacan los realizados por Zmud [2], Ballou o Pazer [1] quienes investigan y profundizan en la forma más efectiva para afrontar la calidad de los datos a través de dimensiones como fiabilidad, consistencia, precisión, oportunidad, etcétera. Otra posibilidad para afrontar la calidad de los datos, como muestran nuevas investigaciones [7] y [9], puede efectuarse también a través de los atributos de los datos.

Por otro lado, existen trabajos relacionados con el tema que muestran el impacto de los errores de los datos sobre la toma de decisiones. El primer trabajo que analizaba este aspecto fue realizado por Ballou y Pazer [3], quienes, sin embargo, no investigaron la interrelación existente entre el tipo del problema para el que era necesario la decisión y la presencia de datos en la información que aportasen el grado de calidad de estos.

Entre los últimos trabajos publicados cabe destacar las investigaciones realizadas por Redman [4] o Klein [5], donde puede observarse como se comienza a aportar información acerca de la calidad de los datos para su utilización en la toma de decisiones.

3. Formas de mejorar la calidad de los datos

Para poder observar la calidad de los datos almacenados es necesario tener alguna medida de calidad que permita evaluar la calidad de la información obtenida. Las informaciones que proporcionan la calidad de los datos almacenados van a formar parte de lo que se conoce como *metadata*. En la actualidad, no existen reglas estrictas de cómo aplicar *metadata*, pero básicamente para conocer la medida de calidad de los datos almacenados pueden realizarse dos tipos diferentes de observaciones.

Por un lado, es posible observar la calidad de la información a nivel de cada atributo individualmente, para lo cual bien se hará por medio del empleo de etiquetas asociadas a los datos, bien se incluirán las dimensiones más

relevantes de la calidad de los datos tales como precisión, oportunidad u otras dimensiones estimadas.

La segunda opción de observación, emplea, a diferencia de la anterior, la tupla completa e incluso, en algunas ocasiones, la tabla relacional en el análisis de la calidad de los datos, considerando los datos como un conjunto global frente al enfoque anterior individual.

De entre estas dos opciones, no se decanta por ninguna de ellas en concreto, sino que se va a seguir una aproximación que es una versión intermedia entre ambas donde cada campo va a tener asociado un cierto nivel de calidad para la toma de decisiones.

En esta aproximación se parte de que dentro de un mismo dominio concreto la calidad de todos los datos es la misma, pero que puede variar si estos mismos datos se encuentran en otros dominios. Esta restricción simplifica significativamente el almacenamiento y recuperación de las informaciones del diccionario de datos, donde se supondrá que estarán almacenadas las informaciones sobre calidad de los datos que se necesitan.

Una de las cuestiones importantes que se plantea es como presentar a los usuarios la información acerca de la calidad de los datos y como almacenar la misma. La respuesta a esta cuestión va a proporcionar diferentes tipos de *metadata* que se le pueden ofrecer a los usuarios.

Como respuesta a esta cuestión se buscará que la presentación de la información al usuario se proporcione bajo un formato que le permita efectuar la toma de decisiones más efectiva. Esto implica que la presentación no sólo debe proporcionar la información sino que se efectuará en un formato acorde a la toma de decisión que el usuario debe realizar.

4. Procesos en la toma de decisiones

Ante lo anteriormente expuesto, se debe exponer que la forma más efectiva para realizar la presentación de la calidad de la información va a estar en función de la estrategia o proceso de toma de decisión que se quiera efectuar.

Se ha escrito mucho durante los últimos años acerca de la problemática de la toma de decisiones, pero en este punto se van a adoptar únicamente dos

aproximaciones, la primera descrita como toma de decisiones conjuntiva y la segunda como toma de decisiones ponderada.

La toma de decisiones conjuntiva toma como punto de partida un conjunto de criterios conocidos y perfectamente especificados y establece que la decisión va a depender de ellos. A cada uno de los criterios se le asocia un nivel mínimo de aceptación que servirá para observar si el dato cumple con el criterio. A la hora de tomar la decisión de que alternativa seleccionar entre las diferentes alternativas existentes, se procede a la evaluación de cada alternativa cotejándola con cada criterio establecido. Las alternativas que no alcancen el nivel mínimo de aceptación en todos los criterios serán desechadas. Por tanto, si una alternativa no cumple con uno sólo de los criterios establecidos, se considerará que ya que no alcanza el nivel mínimo exigido y será catalogada como alternativa no válida. De entre todas las alternativas que hayan superado el nivel mínimo exigido de todos los criterios se tomará la alternativa que los haya cumplido con una mayor holgura.

En cuanto a la toma de decisiones ponderada parte también de una serie de criterios identificados y perfectamente especificados sobre los que va a basar la decisión. Ahora, en esta aproximación, a cada criterio se le asigna una ponderación o peso, que representará la importancia de cada criterio a la hora de efectuar la toma de la decisión. La suma del conjunto de todas las ponderaciones asociadas a cada criterio se corresponderá, normalmente, con uno. Se considera que en este caso se dispone, como en la aproximación anterior, de varias alternativas a seleccionar. Se toma cada alternativa y se evalúa de acuerdo con los criterios establecidos, tomando éstos de uno en uno, obteniendo un resultado que se multiplicará por la ponderación o peso establecido para dicho criterio. Una vez efectuada la multiplicación los valores que se obtienen para cada criterio se suman proporcionando un resultado final para toda la alternativa. La alternativa que se seleccionará de entre todas las existentes será aquella cuyo resultado final sea el mayor de todos.

La principal diferencia entre estas dos aproximaciones se observa en que la aproximación conjuntiva no permite que se compense el nivel mínimo exigido a un criterio con el nivel de otro, sino que obliga a que es necesario que se cumplan todos los criterios, mientras que la aproximación ponderada permite la compensación de valores de los diferentes criterios.

Además, la aproximación ponderada observa a las alternativas desde un punto de vista global, ya que se evalúan todos los criterios, mientras que en la aproximación conjuntiva, el análisis de la alternativa se ve mediante la comparación parcial de cada criterio con un nivel mínimo preestablecido, sin tener en consideración los otros criterios.

Por tanto se puede observar que ambas estrategias seleccionadas ofrecen diferentes puntos de vista en función de sus características a la hora de evaluar una determinada alternativa aunque la aproximación conjuntiva presenta una mayor rigidez.

A la hora de optar por una de las dos aproximaciones, la elección vendrá en función de la relación coste/beneficio, de manera que siempre se busque una minimización de la relación, en la que se tome siempre la mejor decisión con el mínimo coste o esfuerzo.

Por este motivo, el impacto que la calidad de los datos y de la información almacenada tuviera en la toma de decisiones dependerá en gran manera de la complejidad del proceso para el cual se este tomando la decisión. Para conocer la complejidad de un proceso o tarea, esta vendrá dada en función de varias variables como puedan ser el número de alternativas, el número de atributos o el tiempo de apremio.

Incluso considerando que la incorporación de la información referente a la calidad de los datos en una base de datos no presenta dificultades técnicas al diseñador, éste se encuentra con la necesidad de identificar cual va a ser la información apropiada que debe emplear para considerar la calidad de los datos. Esta circunstancia aporta un nuevo problema al diseñador de las bases de datos, ya que deberá de seleccionar la información a emplear y establecer la escala de valores a utilizar.

Se puede considerar que la utilización de una escala de 2 categorías posibles sea más sencillo que si se emplea una escala continua de 100 puntos. También es necesario considerar que si mediante el empleo de esta última escala es posible realizar una toma de decisiones más eficiente, entonces este mayor esfuerzo necesario para tomar la decisión debe ser considerado.

Para comprobar que el suministro de información acerca de la calidad de los datos

proporciona una mejor toma de decisiones, se realizará una toma de decisión partiendo de una misma tarea a la cual se le pasan diferentes muestras de datos en cuanto a presencia de calidad se refiere. Estas muestras de datos oscilarán desde la presencia de información de calidad nula hasta diversos factores que permitirán precisar la calidad de los datos.

5. Planteamiento del experimento

El desarrollo del trabajo experimental es importante para validar y generalizar los resultados obtenidos en la investigación. Este experimento está basado en diversos trabajos presentes en la literatura, en concreto el desarrollado por Pazer [6]. Por tanto, en este artículo se presenta un experimento en el cual se analiza la toma de decisiones a realizar en función de la calidad de los datos disponibles.

El experimento toma como punto de partida para el estudio seis atributos diferentes dispuestos en cinco replicas o alternativas. En cada una de estas réplicas los seis atributos presentan una valoración que será variable en función de los miembros que participan en el análisis de la réplica. El análisis de cada réplica teniendo en cuenta únicamente la valoración, que compondrá una réplica carente de calidad, será confrontado en las hipótesis con los análisis que consideran muestras de datos con calidad. En estas últimas muestras de datos la calidad puede venir determinada por la fiabilidad de los datos, por los pesos de cada atributo e inclusive por ambos aspectos. Los niveles mínimos de tolerancia aceptables para cada criterio, así como la evaluación, han sido especificados para cada réplica y para cada atributo, proporcionando todos estos datos a los evaluadores en ambos experimentos.

El estudio que se presenta con información de calidad de los datos, tiene una solución preferente proporcionando medidas de evaluación objetivas, las cuales se utilizan como base de datos inicial añadiendo una dimensión de calidad en los criterios seleccionados. Al objeto de corregir los errores totales se tiene la adición de pesado, que modifica cada uno de los valores de los atributos mediante la ponderación.

A continuación se presentan las cinco réplicas con sus correspondientes valoraciones para cada atributo, que serán los factores para determinar la toma de decisiones del estudio en cuestión. El sistema para dotar a los atributos de calidad de la

información es proporcionar a estos una fiabilidad, cuyos valores han sido asignados previamente de acuerdo a un rango de 0 a 100 y que quedan descritos en las *Tablas 1*. Igualmente, en estas tablas se muestran los valores de los pesos introducidos al objeto de corregir los errores totales que se presentan en tantos por uno.

Las valoraciones reflejadas en cada una de estas réplicas han sido tomadas partiendo de cinco grupos de trabajo de la Escuela Universitaria de Informática.

Estos grupos de trabajo son independientes entre si y los valores dados por cada uno de los grupos corresponden a las directrices generales dadas por el equipo de investigación para evitar los efectos colaterales de ruido presentes en cualquier toma de datos.

Réplica 1A

Criterios	Fiabilidad	Valoración	Pesos
Caract. A ¹	23	38	0,1
Caract. B ²	15	35	0,2
Caract. C ³	21	24	0,1
Caract. D ⁴	13	31	0,25
Caract. E ⁵	16	26	0,2
Caract. F ⁶	12	10	0,15

Réplica 1B

Criterios	Fiabilidad	Valoración	Pesos
Caract. A	23	35	0,1
Caract. B	15	32	0,2
Caract. C	21	26	0,1
Caract. D	13	30	0,25
Caract. E	16	27	0,2
Caract. F	12	9	0,15

-
- ¹ Piso Céntrico.
 - ² Situado en una zona tranquila.
 - ³ Número de metros cuadrados.
 - ⁴ Luminosidad.
 - ⁵ Cuantía de la renta.
 - ⁶ Antigüedad del edificio.

Réplica 1C

Criterios	Fiabilidad	Valoración	Pesos
Caract. A	23	37	0,1
Caract. B	15	30	0,2
Caract. C	21	26	0,1
Caract. D	13	31	0,25
Caract. E	16	28	0,2
Caract. F	12	11	0,15

Réplica 1D

Criterios	Fiabilidad	Valoración	Pesos
Caract. A	23	36	0,1
Caract. B	15	30	0,2
Caract. C	21	27	0,1
Caract. D	13	29	0,25
Caract. E	16	20	0,2
Caract. F	12	15	0,15

Réplica 1E

Criterios	Fiabilidad	Valoración	Pesos
Caract. A	23	40	0,1
Caract. B	15	30	0,2
Caract. C	21	27	0,1
Caract. D	13	33	0,25
Caract. E	16	22	0,2
Caract. F	12	14	0,15

Tablas 1

5.1. Hipótesis

Idealmente se parte de la premisa que una ayuda en la toma de decisiones debe ser la inclusión de información en la calidad de datos, permitiendo estos datos una mejor toma de decisiones. El conocimiento de la calidad de los datos debe afectar en la toma de decisiones, para determinarlo se diseñan tres hipótesis. De esta forma, si estas hipótesis nulas son rechazadas en el contraste de hipótesis que se realiza frente a las hipótesis alternativas, se tiene que la incorporación de calidad en los datos es de importancia en la toma de decisiones y por tanto será un factor determinante en el diseño de base de datos.

Hipótesis 1 (nula). No tiene influencia en la toma de decisiones la alternativa de incluir un nivel de fiabilidad para cada uno de los atributos que se están analizando.

Hipótesis 1 (alternativa). Tiene influencia en la toma de decisiones la alternativa de incluir un nivel de fiabilidad para cada uno de los atributos que se están analizando.

Hipótesis 2 (nula). No tiene influencia en la toma de decisiones la alternativa de incluir una ponderación para cada uno de los atributos que se están analizando.

Hipótesis 2 (alternativa). Tiene influencia en la toma de decisiones la alternativa de incluir una ponderación para cada uno de los atributos que se están analizando.

Hipótesis 3 (nula). No tiene influencia en la toma de decisiones la alternativa de incluir los factores de fiabilidad y la ponderación para cada uno de los atributos que se están analizando.

Hipótesis 3 (alternativa). Tiene influencia en la toma de decisiones la alternativa de incluir los factores de fiabilidad y la ponderación para cada uno de los atributos que se están analizando.

6. Análisis y conclusiones

Las diferentes valoraciones asignadas en cada análisis al incluir los diferentes formatos de calidad se realizan utilizando un análisis estadístico multivariante [8], que permite comparar los niveles de la multivariabilidad en las tres hipótesis realizadas.

Con este experimento se ha conseguido demostrar la influencia de introducir factores de calidad en la toma de decisiones. El experimento ha sido planteado teniendo en cuenta valores independientes tanto de los pesos como de las fiabilidades presentes en cada una de las características.

En el análisis realizado se plantea el contraste de hipótesis considerando por una parte los datos sin ninguna influencia relativa a la calidad de los mismos y por otra parte teniendo en cuenta la influencia de la fiabilidad y la ponderación asignada a cada una de las características y el conjunto de ambas.

Los resultados obtenidos han demostrado que la introducción de calidad en los datos, es decir, pesos y fiabilidades en la toma de decisiones, prevalece sobre las otras consideraciones. Además, se comprueba que cualquier presencia independientemente de una de ellas prevalece sobre la consideración de los valores de las características tomados sin ninguna aportación relativa a la calidad de los datos. Los resultados obtenidos se presentan en la *Tabla 2*.

ALTERNATIVAS	VALORES κ^2
Atributos con Ponderación	$\kappa^2 = 0.01028$
Atributos con Fiabilidad	$\kappa^2 = 0.01369$
Atributos con Fiabilidad y Ponderación	$\kappa^2 = 0.01432$

Tabla 2

De estos resultados se puede concluir que en los tres casos analizados se rechaza la hipótesis nula frente a la hipótesis alternativa, es decir, la toma de datos con calidad proporciona una mayor fiabilidad en la toma de decisiones que la definición de datos sin información relativa a la calidad de los mismos. Por tanto se demuestra las sinergias existentes entre la calidad de la información y la toma de decisiones.

Del estudio se desprende que se aceptaría con más éxito los datos que están modificados desde su diseño por un coeficiente de fiabilidad frente a los que carecen de ella, y estos frente a los que sólo vienen modificados por la ponderación de cada variable, siendo el caso más favorable aquellas valoraciones en las que se tienen en cuenta tanto la fiabilidad como la ponderación.

Por lo tanto, se puede afirmar que es concluyente que en el diseño de las bases de datos es fundamental la inclusión de los factores de fiabilidad y de ponderación para cada uno de los atributos utilizados de cara a adoptar una correcta toma de decisión.

7. Bibliografía

- [1] D. P. Ballou and H.L. Pazer, Modeling Data and Process Quality Multi-Input Multi-Output Information Systems Management Science, vol 31, nº 2, pp 150-162, 1985
- [2] R. W. Zmud An Empirical Investigation of the Dimensionality of the Concept of Information Decision Science, vol 9, pp187-195, 1978
- [3] D. P. Ballou and H.L. Pazer. Framework for the Analysis of Error in Conjunctive, Multi-Criteria, Satisficing Decision Processes Decision Science, vol 21, nº 4, pp752-770, 1990
- [4] T. C. Redman Data Quality for the Information Age Boston:Artech House, 1996
- [5] B. D. Klein, D. L. Goodhue, G. B. Davis

Can Humans Detect Errors in Data? Impact of Base Rates, Incentives and Goals
MIS Quarterly, vol 21, pp 169-194, 1997

[6] I. N. Chengalur-Smith, D. P. Ballou, H. L. Pazer
The Impact of data Quality Information on Decision Making: An Exploratory Analysis
IEEE transactions on Knowledge and Data Engineering, vol 11, n° 6, 1999

[7] P. K. Kapur, R. B. Garg, S. Kumar
Contributions to Hardware and Software Reliability
World Scientific Pub Co, 1999

[8] D. C. Montgomery
Introduction to Statistical Quality Control. Third Edition
John Wiley and Sons, Inc. 1997

[9] R. S. Kenett, E. R. Baker
Software Process Quality: Management and Control
Marcel Dekker, Inc. 1999