

ZigBee Pulse Oximeter

J.M. Castillo-Secilla*, Joaquín Olivares*, J.M. Palomares*, J.M. Soto-Hidalgo*, J.C. Gamez*, Lilia D. Tapia*

*Dept. of Computer Architecture, Electronics, and Electronic Technology.

University of Córdoba.

e-mail: jmcastillo@uco.es, olivares@uco.es, jmpalomares@uco.es,
jmsoto@uco.es, jcgamez@uco.es, ltapia@uco.es

Abstract—This work presents a prototype to adapt a standard pulse oximeter by turning it into a wireless device using ZigBee. Patient's data are extracted and transmitted to the server in real time through a Wireless Sensor Network. This Wireless Sensor Network is deployed using the mesh topology in order to reach the maximum reliability in the communications. The pulse oximeter is based on a Nellcor DS-100a probe and is controlled by an Arduino FIO with a XBee wireless modem. The amplifier circuit which is designed to extract the information of the pulse oximeter probe is included in this work.

Index Terms—code

ZigBee, Xbee, Wireless Sensor Network, Pulse Oximeter, e-Health, Telemetry in health care

I. INTRODUCTION

Nowadays, health monitoring systems are one of the most important research areas in the scientific field. This work is focused in the monitoring systems for hospital applications. In hospitals, life constants of the patients are monitored by several systems, most of which use sensors that are wire-connected to the monitoring devices. The usage of wires makes them uncomfortable and there is a lack of mobility both for the patient and for the hospital staff. This work presents the design of a wireless pulse oximeter which is able to store the information about the patient in real-time through a wireless sensor network. This information is stored in a Database server, allowing the query for all data from the patients. This queries will allow to derive new information about the health status of the patients.

Wireless telemetry [Jovanov(2005)], [Jovanov(2000)], [Hameed(2003)] monitoring will improve patient mobility. Some of the aims of the wireless telemetry are to increase the patient's satisfaction and to improve working conditions for the medical team. Nowadays, there are some wireless telemetry devices in the market. Several Bluetooth communication wireless pulse oximeters [Morón et al.(2005)Morón, Luque, Casilari, and Gázquez] are commercially available. In this work, three different wireless communication protocols (Wifi, Bluetooth and ZigBee) have been studied. A Wireless Sensor Network to monitor patients has been designed.

The rest of the paper is organized as follows. Section II describes the theoretical principles of the pulse oximeter. In Section III, the wireless communication is described. Section IV describes the microcontroller used to process the pulse and SpO₂ values. In Section V, the pulse oximeter probe and its components are detailed. Section VI describes the architecture of the proposed system. Section VII describes the

results obtained by monitoring patients in hospital and finally, Section VIII presents the main conclusions of this work.

II. PRINCIPLES OF OXIMETRY

In the decade of the 1860s, it was discovered that the colored substance in blood, hemoglobin, was also the carrier of oxygen. At the same time, it was noticed that the absorption of visible light by a hemoglobin solution varies with oxygenation. This is because the two common forms of the molecule, oxidized hemoglobin (HbO_2) and reduced hemoglobin (Hb), have significantly different optical spectra in the wavelength range from 500nm to 1000nm [Towsend(2001)].

Based on the Beer-Lambert law [Webster(1997)], pulse oximetry measures the effect of arterial blood on the tissues taking into account the intensity of the transmitted light [Cheung et al.(1989)Cheung, Gautlitz, Mason, Prosser, Smith, Wagner, and Hunsaker]. The volume of blood in the tissue is proportional to the arterial pulse. The highest blood volume occurs during systole, whilst the lowest one is presented in diastole. Blood absorbs part of the light that passes through the tissue. The emergent light is proportional to the volume of blood present in the tissue. This light is modified with each heartbeat. Besides, the absorbance coefficient of oxyhemoglobin is different from that of deoxygenated hemoglobin for most wavelengths of light. Eq. 1 states the formula to compare both types of hemoglobin.

$$\%S_aO_2 = \frac{[HbO_2]}{[Hb] + [HbO_2]} \times 100\% \quad (1)$$

A pulse oximeter shines light of two wavelengths through a tissue bed such as the finger or the earlobe, and measures the transmitted light signal. The device operates on the following principles:

- The light absorbance of oxygenated hemoglobin and deoxygenated hemoglobin at two wavelengths is different.
- The pulsatile nature of arterial blood results in a waveform on the transmitted signal that allows the absorbance effects of arterial blood to be identified from those of non pulsatile venous blood and other body issue. By using a quotient of the two effects at different wavelengths it is possible to obtain a measure requiring no absolute calibration with respect to overall tissue absorbance. This is a clear advance of pulse oximeters over previous types of oximeters.

	ZigBee®	Wi-Fi™	BT™
	802.15.4	802.11b	802.15.1
Transmission Range (m)	1 – 100	1 – 100	1 – 100
Battery Lifetime (days)	100 – 1000	0.5 – 5	1 – 7
Network Size (# nodes)	> 65000	32	7
Stack Size (Kbytes)	4 – 32	1000	250
Transference Rate (Kb/s)	20 – 250	11000	720

TABLE I
COMPARATIVE BETWEEN TECHNOLOGIES

- With adequate light, scattering in blood and tissue will illuminate sufficient arterial blood, allowing a reliable detection of the pulsatile signal. The scattering effect requires the empirical calibration of the pulse oximeter. On the other hand, this effect allows a transmittance path around the bone in the finger.

As it has been previously mentioned, a pulse oximeter shines light of two wavelengths, red and infrared, and each light signal has an AC and DC component. Both AC and DC from each wavelength have to be measured [Watthanawisuth et al.(2010)Watthanawisuth, Lomas, Wisitsoraat, and Tuantrant] using Eq. 2.

$$R = \frac{\log\left(\frac{R_{AC}}{R_{DC}}\right)}{\log\left(\frac{I_{R_{AC}}}{I_{R_{DC}}}\right)} \quad (2)$$

R is the comparison value between the red and infrared lights, which allows to calculate the SpO₂ value from the calibrated Eq. 3.

$$SpO_2 = 10.0002R^3 - 52.887R^2 + 26.817R + 28.293 \quad (3)$$

The heart rate is determined by measuring the time between maximum values in the infrared light. The heart rate is calculated using Eq. 4.

$$BPM = \frac{60}{Period(secs)} \quad (4)$$

III. WIRELESS COMMUNICATION

In order to get rid of data wires in hospital, wireless communication will be used. There are several wireless technologies available, which are compared in Table I. A wireless pulse oximeter needs certain characteristics as low power consumption, high transmission range and easy integration (both physical and functional). These facts makes ZigBee as the best protocol to ensure these requirements. Besides, one of the main applications of ZigBee [Alliance(2010)] is e-Health. These technologies ensure low power consumption and highly reliable communications.

A wireless XBee™ module by MaxStream® has been used to transmit the acquired signal. This XBee module is based on the ZigBee protocol.

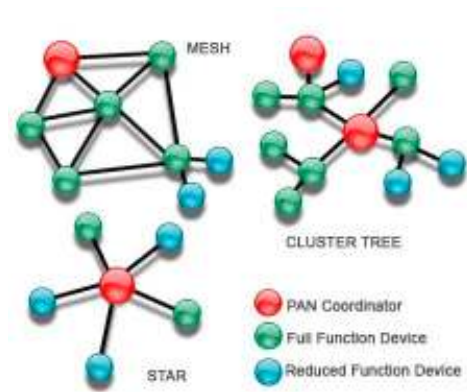


Fig. 1. Data Redundancy Vs Hardware Redundancy

A. ZigBee Devices

ZigBee allows up to 254 nodes per network. However, this protocol allows to make groups with up to 255 nodes per cluster, and thus, to reach a higher limit of 64770 nodes in the same network. There are three types of devices in ZigBee:

Coordinator: The most capable device, the coordinator forms the root of the network tree. It can act as a bridge to other networks.

Router: This type of device is able of behaving as an intermediate router, passing on data from other devices determining the best path.

End Device: It contains just the minimal functionality to communicate to the parent node (either the coordinator or a router); it cannot relay data from other devices. This relationship allows the node to be asleep for a significant amount of the time, thereby providing a long battery life.

According to their functionality, these devices are grouped into:

FFD: Also known as active node. Using additional memory and computing capacity, they can act as Coordinator or ZigBee Router. They can also behave as an interface with final users.

RFD: Also known as passive node. They have limited capacity and functionality, so that they achieve great simplicity at a low cost. Basically, they are used to place sensors/actuators within the network.

B. ZigBee Topologies

As shown in Figure 1, ZigBee supports three network topologies:

The **Star Topology** has the following advantages: easy synchronization, low power modes, and low latency. On the other hand, the main disadvantage is the centralization of the coordinator, which could provoke the stop of the entire network.

The advantages of the **Tree Topology** are: low routing cost, SuperFrames that allow the use of sleep modes, and MultiHop communication. However, the recuperation costs are very expensive when a route crashes.

The **Mesh Topology** has these advantages: robust multihop communication, the network is more flexible to keep up with structure changes, and lower latency than star and tree topologies. The main drawbacks are: the route discovery is more expensive, and it needs additional storage space to save routing tables.

Mesh Topology has been selected in this work. Hospitals require robust networks that automatically recalculate new routes when the established routes fall, in order to prevent a loss of information. It needs to ensure that all packets arrive at their destinations and with real-time constraints, in this sense, the mesh topology is the appropriate one.

IV. MICROCONTROLLER: ARDUINO FIO

Once the pulse oximeter algorithm is known and the wireless protocol has been determined, the following step is to select the microcontroller. It will digitalize the patient's signal and will send it to the destination. Several microcontrollers with XBee module slots are available, such as Rabbit RCM4510W, Arduino FIO, Arduino NG with XBee Shield, etc.

In a first approximation, Rabbit RCM4510W was selected to work with. A preliminary study over this board was needed to determine if it could be programmed as a pulse oximeter. In a previous work [Castillo-Secilla et al.(2010)Castillo-Secilla, Bellido, and Olivares], a study of the RCM4510W in terms of energy consumption and hardware reliability was developed. The conclusions of this work determined that real power consumption does not follow the expected response provided by Rabbit manufacturer. Besides, the ADC (Analog Digital Converter) implemented in RCM4510W RabbitCore® is very slow to acquire all the required data in order to reconstruct the pulse oximeter signal. For these reasons, Rabbit was discarded.

On the other hand, Arduino is an open-source electronics prototyping platform based on flexible, easy-to-use hardware and software. Arduino can sense the environment by receiving an input from a variety of sensors and can affect its surroundings by controlling lights, motors, and other actuators. The microcontroller on the board is programmed using the Arduino programming language and the Arduino development environment. Arduino projects can be stand-alone or they can communicate with running software on a computer.

The Arduino FIO is a microcontroller board based on the ATmega328P. It has 14 digital input/output pins, 8 analog inputs, an on-board resonator, a reset button, and holes for mounting pin headers. It has a Lithium Polymer battery and can be charged with a USB cable. An XBee socket is available on the bottom of the board.

All these features, combined with its small size (2.8cm × 6.6cm), makes Arduino FIO suitable for pulse oximeter design along with patient comfort.

V. NELLCOR PULSE OXIMETER PROBE

The main goal of this work is to suppress the wires of pulse oximeters while keeping the budget as low as possible. For these reasons, it is very important to select a commonly

Microcontroller	ATmega328P
Operating Voltage	3.3V
Input Voltage	3.35 -12 V
Input Voltage for Charge	3.7 - 7 V
Digital I/O Pins	14 (of which 6 provide PWM output)
Analog Input Pins	8
DC Current per I/O Pin	40 mA
Flash Memory	32 KB
SRAM	2 KB
EEPROM	1 KB
Clock Speed	8MHz

TABLE II
ARDUINO FIO CHARACTERISTICS



Fig. 2. Nellcor DS100-a probe

used pulse oximeter probe, which will allow the reduction of deployment costs of a new wireless pulse oximeter. There are several pulse oximeter probes in the market based in the Beer-Lambert law (Nellcor, Ohmeda, etc). The selected probe for this work is the well known Nellcor DS-100A. This probe is widely used and can be connected to the designed wireless pulse oximeter with total compatibility. Besides, it is a tested device in hospitals, so its use is expected to be welcomed by the medical community.

A. LED

One of the large improvements of modern pulse oximeters over earlier oximeters is the use of LEDs as their light source. LEDs can transmit large intensities of light proportional to the amount of drive current. As there are two LEDs in the system (red and infrared), the microcontroller is responsible for their timing control. The timing of pulsations is critical because the photodiode cannot distinguish between different light wavelengths. Pulse oximeter relies on the microprocessor system to synchronize the pulsations of the LEDs with the samples taken by the ADC. This highly precise synchronization will allow to properly determine which LED is the transmitter of every detected absorbance value by the photodiode.

The two chosen wavelengths for pulse oximetry are 660

Pin	Function
2	Diode Red Cathode//Diode Infrared Anode
3	Diode InfraRed Cathode//Diode Red Anode
5	Ground for Phototransistor
9	Circuit input for Phototransistor

TABLE III
NELLCOR DS-100A PINOUT



Fig. 3. Nellcor DS-100a Connector

nm and 940 nm. They were chosen because of the availability of LEDs at these wavelengths. Besides, theory states that the extinction coefficients, Hb and HbO₂, of each light should be selected to vary as much as possible. In this sense, HbO₂ has been determined to have a higher extinction coefficient than Hb at 940 nm and a lower extinction coefficient at 660 nm. In other words, as SaO₂ increases, the absorbance light increases at 940 nm and decreases at 660 nm.

B. Phototransistor

The photodetector is a silicon phototransistor that produces a voltage which is linearly proportional to the striking light intensity. Advances in silicon technology allow the phototransistor to be small enough to fit in reduced-size, finger tip probes. These advances have made pulse oximeters much more accurate and convenient than earlier devices. Previous oximeter devices needed frequent calibration because the photoelectric devices, used as sensors, were often inconsistent [Miller(1966)]. A photodiode cannot distinguish between a red and an infrared light, but to deal with it, the microprocessor system alternatively turns each LED on and off. The pulse oximeter takes samples of the photodiode output while the red LED is on, while the infrared LED is on, and while both are off. By sampling with both LEDs off, the pulse oximeter is able to subtract any ambient light that may be present [Pologe(1987)].

1) *Tests: Nellcor DS-100a:* Finally, the Nellcor probe has an open collector phototransistor and its pinout is described in Table III and Fig. 3.

The signal given by the probe is about 0.1-1mV, which is too low to be directly processed by the microcontroller. Therefore, a circuit to amplify and process the signal is required [Webster(1997)], [Baran et al.(2006)Baran, Chen, Stanford, and Yamarkovich]. In the following section, the necessary circuit to filter and amplify the signal in order to digitalize it with the microcontroller has been developed.

VI. SYSTEM ARCHITECTURE

Nowadays, monitoring systems in hospitals are usually designed using wires. For pulse oximeters, there exist some wireless devices, although they are not able to set up an

entire network which allows to manage all the system within a server. The main advantage of a server is the possibility of including Artificial Intelligence (AI) in the system, which would provide the capacity of performing complex inferences. Thanks to this AI, the system could not only act as a mere data storage about patients, but it could also study each case in real-time, sending alerts to the medical staff when the state of any patient changes. In this work, a first approximation composed of a wireless sensor network and a database server has been designed and implemented.

In Fig. 4, the architecture of this system is described. In the following sections, the design of this system will be explained, focusing on the used probe, the wireless technology, the microcontroller, the main server of the system, and the database server.

A. Operational Amplifiers and Circuits

In this first approximation, operational amplifiers by a $\pm 12V$ source to amplify and filter the signal of the probe have been used. The circuit has been designed in four stages to keep the different recommendations described in the literature [Webster(1997)], [Cysewska-Sobusiak(1995)], [Graeme(1992)], [Burr-Brown Corp.(1994a)], [Burr-Brown Corp.(1994b)], [Burr-Brown Corp.(1994c)] in mind.

1) *Filter 1: Filtering 60Hz Noise:* In this first stage, the signal passes through a bandpass filter to eliminate DC offset and to filter 60Hz noise. Pin 9 of the Nellcor DS-100a is connected to the amplifier circuit. This is an inverting bandpass filter with calculations. A High Pass Filter is created using R1 and C1, whilst C2 and R4 controls the construction of a Low Pass Filter. The output of this circuit is connected to the input of the second stage filter.

2) *Filter 2: Setting the Cutoff Frequency to 0.5Hz :* By setting the cutoff frequency to 0.5Hz, the noise of the original signal is reduced. In order to do this, a new High Pass Filter with a cutoff frequency of 0.5Hz [Webster(1997)] has been developed.

$$f_c = \frac{1}{2\pi \cdot R_6 \cdot C_3} = \frac{1}{2\pi \cdot 330k \cdot 1\mu F} = \frac{1}{2.07} = 0.4822Hz \quad (5)$$

3) *Filter 3: Amplifying the signal:* As required by the used microcontroller, the signal is amplified to 3.3V. The R9 and R10 resistances provide the main gain of the circuit. Several tests have been performed to measure the heartbeat and oxygen values, and the best configuration to read them in the microcontroller is obtained by setting R9 to 38.9M Ω and R10 to 100K Ω . With them, a gain value of 398 is obtained. Therefore, the signal has been amplified from 8mV to 3.3V, approximately.

$$G = \frac{R_9}{R_{10}} = \frac{38.9M\Omega}{100K\Omega} = 398 \quad (6)$$

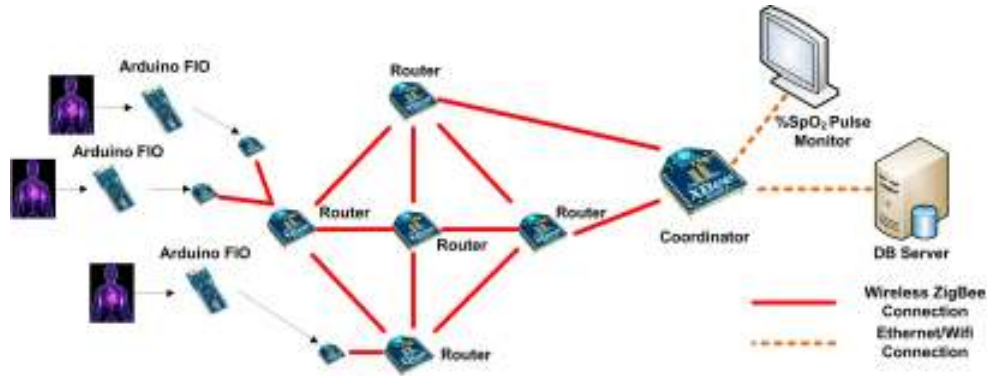


Fig. 4. System Architecture

4) *Filter 4: Setting the Cutoff Frequency to 5Hz* : The final stage involves cutting out the low frequencies produced by the motion of the finger. A sallen-key Low Pass Filter with 5Hz cutoff has been designed [Webster(1997)]. The output of this stage is connected to an analog input of the microcontroller.

$$f_c = \frac{1}{2\pi\sqrt{R_{11}R_{12}C_4C_5}} = \frac{1}{2\pi\sqrt{100k \cdot 100k \cdot 470nF \cdot 220nF}} = 4.9494Hz \quad (7)$$

B. Xbee USB Coordinator

The coordinator is the most important element of a ZigBee network and, therefore, there should be one in a network. The coordinator is responsible for monitoring both the whole network and the communication paths to be followed by devices in order to connect with each other. The XBee USB is developed by Digi. The device has the configuration shown in Table IV.

A Java library, Xbee-api.lib [xbe(2010)], has been used to extract the information from the coordinator. This library contains all the required classes to set up a communication interface between the XBee module and the ZigBee wireless network gateway. In this work, the XBee USB coordinator has been connected to serial port COM3 at 9600 bauds data rate. The software in the coordinator has been developed to receive data from the probe and to send it to the Database server. In this stage of the work, the information is sent from the coordinator to the Database Server using an ethernet connection. In a future, when this system will be deployed in hospitals, it could be implemented using a Wifi network to send the information between the coordinator and the Database Server.

C. Database Server

The Database Server is implemented using MySQL Server in a Linux machine. This database stores the following information about patients: patient name, gender, age, historical pulse, historical % SpO₂, and drug treatment. This Database

ZigBee Item	Value
PAN ID	A3321E3C4B9F63F
SD (Channel Scan)	3
ZS (ZigBee Stack)	0
SH (High Address)	13A200
SL (Low Address)	400A00D6
NI (Node Identifier)	COORDINATOR
BH (Broadcast Radio)	0
AP (API Mode)	2
AO (API Output Mode)	0

TABLE IV
MAIN CONFIGURATION IN ZIGBEE COORDINATOR.
SEE [ALLIANCE(2010)] FOR FURTHER DETAILS.

design has been developed to keep up with the different experiments of this work, but it could be easily modified to add more fields, if necessary.

In a future, using this database, it will be possible to process information about patients using Artificial Intelligent techniques in order to obtain new knowledge. This technique will provide some advantages, such as: real-time automatic determination whether a patient suffers any cardiac or respiratory problem. In this case, an alarm will be generated in order to inform to the staff. Besides, new medical research will be possible: comparing multiple histories to study different population responses; responses after being treated with certain drugs, etc.

VII. RESULTS

In order to check the reliability of the network and the developed pulse oximeter, various tests were performed. The first type of experiments, called Packet tests, tried to confirm that every packet reached its destination, even if some paths were unavailable. The second experiment, called Measurement test, checked the validity of the obtained values (pulse and SpO₂). And the final experiment tested the network with several pulse oximeters running and it was called pulse oximeter coexisting test.

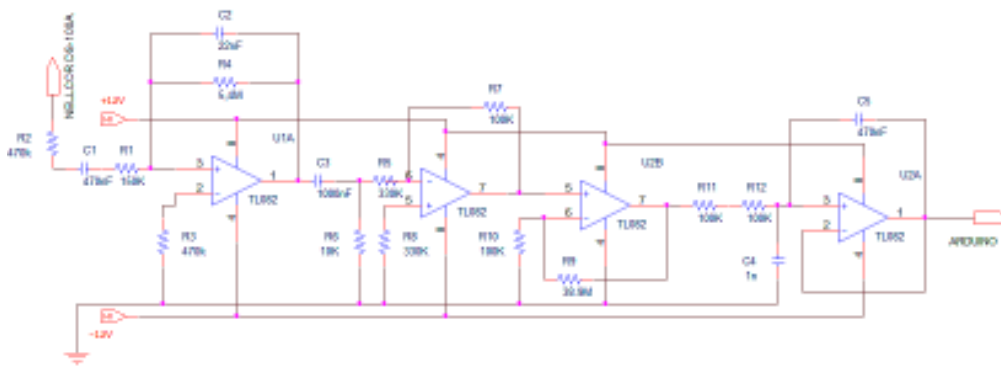


Fig. 5. Amplifier Circuit

A. Packet Tests

To check the communications reliability, a program to count the number of received packets in the coordinator was developed. This test was launched in a controlled real-life environment:

- 3 Rooms divided by walls.
- 5 Router devices.
- 3 End devices.
- Electromagnetic contamination environment with Bluetooth and Wifi signals.

Using the previously presented configuration, the amount of packets sent by each end device and received by the coordinator was 100%. The test was performed during 30 minutes using a pulse/SpO2 simulator connected to the probe.

The second packet test consisted in the elimination of the direct connection between an end device associated to one pulse oximeter and its parent node acting as a router. This test tried to check if the end device was able to reconnect to another router. This test was successful, and the statement of Casilari et al. has been confirmed [Casilari(2010)]: “The operation to reconnect any end device to an FFD takes over 2 seconds”. Obviously, during the reconnection time, the node is unable to transmit any information to the coordinator. Nevertheless, the end device has been programmed to store the information in a buffer in order to avoid the loss of data. When the node re-associates itself to another router, it sends all stored information.

Sometimes, if an end device is not able to reconnect to another router, it may connect directly to the coordinator, if there exists enough coverage between both end points. This situation reduces the amount of traveling time for packets, although it makes the coordinator to be more overloaded.

B. Measurement Tests

In order to test that the data obtained with the developed pulse oximeter are correct, a comparison between sampled data by the designed system and by an ISSO certified (to use in hospitals) professional pulse oximeter was carried out.

In the experiment, two probes were used. The first one was connected to the ISSO pulse oximeter, and the other one was

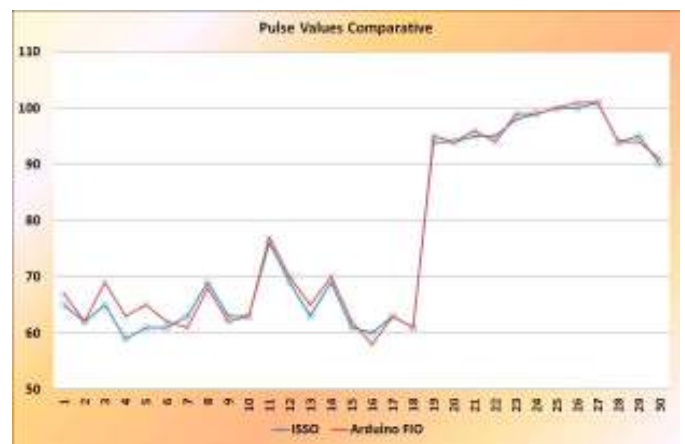


Fig. 6. Heartbeats Results



Fig. 7. O2 Saturation Results

connected to the Arduino pulse oximeter.

Fig. 6 shows several tests with more than 35 values of pulse in different times. Results prove that ISSO certified pulse oximeter and Arduino based pulse oximeter values are similar and differences between both devices are about ± 1 pulses. These values can be considered to be almost equal, even though both pulse oximeters are completely different regarding both algorithm and hardware.

Analyzing the SpO2 values in Fig. 7, it can be observed that both the ISSO pulse oximeter and the Arduino based pulse oximeter differ in less than 1% of the SpO2 value.

These results demonstrate that the Arduino based pulse oximeter has a similar response to the ISSO certified pulse oximeter in both SpO2 and pulse measurements.

C. Pulse Oximeter Coexisting Tests

In this test, 3 pulse oximeters were running simultaneously with successful results. During all the running time, the extracted data of each person reached in real-time the coordinator. In a future, tests in a hospital will be necessary to adjust the network to use a greater number of pulse oximeters.

VIII. CONCLUSIONS

A real-time wireless sensor network for health monitoring using ZigBee protocol is proposed. A pulse oximeter using an Arduino FIO and a Nellcor DS-100a have been designed in order to monitorize patients in hospitals. The pulse and SpO₂ of every patient is collected and sent through the wireless network to the central server. This central server manages the network and sends the information to a Database Server.

Results in terms of reliability and technological viability have been proved to be successful and they ensure that this system could be deployed in any hospital. This system has the capability of storing information and performing advanced inferences to extract new knowledge about the patient state.

In a near future, authors are considering the possibility of including new medical devices into the wireless network using ZigBee protocol. This work will be the basis foundation for the wireless adaptation of those devices.

REFERENCES

- [Jovanov(2005)] E. Jovanov, Wireless technology and system integration in body area networks for m-Health applications, in: Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, volume 7 VOLS, pp. 7158–7160.
- [Jovanov(2000)] E. Jovanov, Wireless personal area networks in telemedical environment, Proceedings of the IEEE EMBS International Conference on Information Technology Applications in Biomedicine (2000) pp.74–78.
- [Hameed(2003)] K. Hameed, The application of mobile computing and technology to health care services, Telematics and Informatics Vol. 20, Iss. 2 (2003) pp. 99–106.
- [Morón et al.(2005)Morón, Luque, Casilari, and Gázquez] M. Morón, J. Luque, E. Casilari, J. Gázquez, A wireless monitoring system for pulse-oximetry sensors, in: ICW '05 Proceedings of the 2005 Systems Communications, pp. 79–84.
- [Towsend(2001)] D. N. Towsend, Pulse oximetry, Medical Electronics (2001) pp. 32–42.
- [Webster(1997)] J. Webster, Design of pulse oximeters, ISBN 0750304677, IOP Publishing Ltd, 1^o edition, 1997.
- [Cheung et al.(1989)Cheung, Gautlitz, Mason, Prosser, Smith, Wagner, and Hunsaker] P. Cheung, K. Gautlitz, L. Mason, S. Prosser, R. Smith, D. Wagner, S. Hunsaker, Feedback-controlled method and apparatus for processing signals and used in oximetry, US patent 4,819,646 (1989).
- [Watthanawisuth et al.(2010)Watthanawisuth, Lomas, Wisitsoraat, and Tuantranont] N. Watthanawisuth, T. Lomas, A. Wisitsoraat, A. Tuantranont, Wireless wearable pulse oximeter for health monitoring using ZigBee wireless sensor network, Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on (2010) 575 – 579.
- [Alliance(2010)] Z. Alliance, <http://www.zigbee.org/>, Last Access October, 2010.
- [Castillo-Secilla et al.(2010)Castillo-Secilla, Bellido, and Olivares] J. Castillo-Secilla, F. Bellido, J. Olivares, Experimental procedure for the characterization of the power consumption and reliability in ZigBee mesh networks, 2010 Third International Conference on Advances Mesh Networks (2010) pp. 13–17.
- [Miller(1966)] S. Miller, A textbook of clinical pathology, Baltimore, MD: Williams and Wilkins, 1966.
- [Pologe(1987)] J. Pologe, Pulse oximetry: technical aspects of machine-design, Int. Anesthesiol 25 (1987) 137–153.
- [Baran et al.(2006)Baran, Chen, Stanford, and Yamarkovich] J. Baran, K. Chen, W. Stanford, M. Yamarkovich, Pulse transmit measuring device, 2006.
- [Cysewska-Sobusiak(1995)] A. Cysewska-Sobusiak, Problems of processing reliability in noninvasive measurements of blood oxygen saturation, Optoelectronic and electronic sensors SPIE 2634 (1995) pp. 163–171.
- [Graeme(1992)] J. Graeme, Phase compensation optimizes photodiode bandwidth, EDN 7 (1992) pp. 177–184.
- [Burr-Brown Corp.(1994a)] Burr-Brown Corp., Opt101: Data sheets: Monolithic photodiode and single-supply transimpedance amplifier (1994a).
- [Burr-Brown Corp.(1994b)] Burr-Brown Corp., Application bulletin ab-075: Photodiode monitoring with op amps (1994b).
- [Burr-Brown Corp.(1994c)] Burr-Brown Corp., Application bulletin ab-077: Designing photodiode amplifier circuits with opa128 (1994c).
- [xbee(2010)] A Java API for Digi XBee/XBee-pro OEM RF modules. <http://code.google.com/p/xbee-api/>, 2010, 2010. Last Access October 2010.
- [Casilari(2010)] E. Casilari, Modeling of current consumption in 802.15.4/ZigBee sensor motes, Sensors 2010 10 (2010) pp. 5443–5468.

El resumen lingüístico como herramienta para el apoyo a la toma de decisiones

Rita Castillo-Ortega, Nicolás Marín, Carlos Molina, Daniel Sánchez

Dpto. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Granada, España

E-mail: {rita,nicm}@decsai.ugr.es

Dpto. de Informática, Universidad de Jaén, Jaén, España

E-mail: carlosmo@ujaen.es

European Centre for Soft Computing, Mieres, Asturias, España

E-mail: daniel.sanchezf@softcomputing.es

Abstract—En una sociedad en la que el conocimiento es esencial nos surge la necesidad de manejar adecuadamente ingentes cantidades de datos. Lo podemos ver en nuestra vida diaria, pero cobra una mayor importancia en el ámbito empresarial. Empresas y organizaciones generan y consumen grandes volúmenes de datos con el fin de llevar a cabo sus actividades. Pero los datos en sí no tienen un impacto en el desempeño de las actividades si no somos capaces de obtener información a partir de ellos que al mismo tiempo sea útil e inteligible para nosotros. La posibilidad de generar resúmenes lingüísticos de series de datos en el ámbito empresarial se presenta como una poderosa herramienta que facilitará el proceso de toma de decisiones.

Index Terms—Resumen Lingüístico, Series de datos temporales, Modelo de datos multidimensional, Business Intelligence, Soporte a la toma de decisiones.

I. INTRODUCTION

La capacidad de manejar grandes volúmenes de datos se perfila cada vez más necesaria en una sociedad que sin duda alguna está basada en el conocimiento. El importante número de grandes empresas, así como organizaciones u organismos públicos, que generan y consumen ingentes cantidades de datos con el fin de llevar a cabo sus actividades, son un buen ejemplo de ello. La mayoría de estos datos están relacionados con la dimensión temporal de una u otra manera.

Pero no sólo el manejo de los datos es útil; del mismo modo, el proceso que permite realizar la extracción de la información a partir de conjuntos de datos, se está volviendo cada vez más importante para nuestro entorno. La importancia de este proceso se debe al hecho de que permite a los usuarios realizar tareas tan importantes como el análisis en la toma de decisiones, pronóstico o predicción [2].

Como se ha comentado anteriormente, la posibilidad de realización de pronóstico y predicción basados en el estudio de las series de datos es muy importante. En este trabajo, nosotros nos vamos a centrar en otra utilidad como es la descripción de la información que alberga la serie temporal. Existen diversas formas de presentar la información que se obtiene a través del análisis de las series de datos. En particular, en nuestro caso nos interesa que la presentación de esta información se realice en forma de resumen.

Los receptores de los resultados obtenidos son los usuarios con formación experta o no, de modo que es muy conveniente

realizar un resumen de la información usando el lenguaje natural. Este tipo de resúmenes reciben el nombre de resúmenes lingüísticos y gracias a ellos es posible ofrecer a los usuarios una información más comprensible y que al mismo tiempo les sea de más utilidad a la hora de tomar determinadas decisiones.

En este trabajo se presentan algunos de los enfoques que han sido desarrollados con el fin de realizar resumen lingüístico sobre las series de datos temporales obtenidas de almacenes de datos. Dichos enfoques usan como herramientas para resolver este problema conceptos de *conjuntos difusos* (FS - Fuzzy Sets) [28]. En la literatura se pueden encontrar diversos enfoques basados en el uso de *Soft Computing* para la realización de resúmenes de datos. De entre estas propuestas, el resumen lingüístico posee un especial interés debido a la ya mencionada necesidad de obtener sentencias cercanas al lenguaje natural al describir grandes conjuntos de datos.

El resumen lingüístico de series de datos temporales es de mucha utilidad en el ámbito referente a *Sistemas de Información* (IS - Information Systems), debido a la importancia de la dimensión temporal en el desempeño del negocio. Las soluciones basadas en *Business Intelligence* permiten a los administradores o gerentes de empresas, organizaciones u organismos, obtener un conocimiento más preciso del proceso que desempeñan y las operaciones comerciales o sociales que llevan a cabo, con el fin de apoyar una mejor toma de decisiones. En general se puede esperar que las compañías que hacen uso de algún tipo de tecnología para facilitar la toma de decisiones tengan un mayor rendimiento que las compañías que precinden de ella. Una parte muy importante de las herramientas relacionadas con Business Intelligence está basada en el uso del modelo de datos multidimensional y las operaciones OLAP (OnLine Analytical Processing) llevadas a cabo sobre él y que permiten la consulta de grandes cantidades de datos [23].

El modelo de datos multidimensional es un modelo ampliamente extendido que se basa en el uso de los llamados cubos de datos (también conocidos como hipercubos o data cubos) los cuales están orientados al análisis de datos. Cada cubo de datos almacena una colección de hechos numéricos, llamados medidas, que se encuentran descritos por un conjunto de dimensiones. El cubo alberga en cada una de sus celdas los datos relacionados con los elementos, agregados en cada una

de las dimensiones.

Los cubos de datos, como norma general, contienen una dimensión *tiempo* debido al importante papel de ésta en general y referente a la actividad comercial en particular. Muchas de las operaciones OLAP que se aplican sobre la dimensión temporal de los cubos de datos producen series de datos temporales. Un gran número de autores han dedicado sus esfuerzos e investigación en realizar minería de datos (DM - Data Mining) sobre estas series temporales. En [1], los autores presentan una extensa visión general de algunos enfoques a este respecto que son de interés.

Cada vez son más las voces que relacionan las tareas llevadas a cabo en los procesos de Data Mining con el soporte a la toma de decisiones [21]. En el presente trabajo consideramos el resumen lingüístico de datos como un caso particular de Data Mining ya que la entrada es la misma y como resultado obtenemos información no trivial, novedosa y potencialmente útil que antes no se conocía, pero con la cualidad añadida de que las salidas se realizan usando lenguaje natural.

A continuación se presenta la estructura del trabajo: La sección II está dedicada a establecer algunos conceptos básicos sobre el resumen lingüístico. En III se aborda la importancia de la dimensión temporal en nuestro entorno y la necesidad del estudio de series temporales. En la sección IV retomamos conceptos anteriores para centrarnos en el resumen lingüístico de datos con técnicas difusas. Nuestro enfoque en la realización de resúmenes lingüísticos de datos temporales así como una plataforma para la obtención de información relevante a partir de grandes cantidades de datos es presentada en la sección V. Finalmente en la sección VI se muestran algunas conclusiones relativas al trabajo.

II. EL RESUMEN LINGÜÍSTICO

De acuerdo con el diccionario de la Real Academia Española de la Lengua, *resumen* es una “exposición resumida en un asunto o materia”. Se trata también de la “acción y efecto de resumir o resumirse”. Y lo que hacemos al *resumir* es “reducir a términos breves y precisos”, o, “considerar tan sólo y repetir abreviadamente lo esencial de un asunto o materia”.

Siguiendo las ideas e investigaciones de Pinto [13], diremos que el “resumen es el documento referencial más completo y por consiguiente el que mejor representa la información original, ofreciendo una visión global del contenido del documento”.

El hecho de que un resumen sea *breve* no implica que en él se refleje lo *esencial* acerca de algo, aunque esta situación sería la más deseable. En muchas ocasiones estas dos palabras son usadas como sinónimos cuando en realidad no lo son. Es cierto que existen diversas circunstancias o situaciones en las que no es necesaria tanta precisión al hablar o definir términos, ya que existen patrones de conocimiento común respecto a un cierto tema, o que son compartidos por una determinada comunidad. Debemos de tener muy en cuenta las diferencias existentes entre los términos *breve* y *esencial*.

La situación ideal a la hora de enfrentarnos a la realización de un resumen sería que fuésemos capaces de ofrecer la

información esencial de una forma breve y concisa. Pero en la mayoría de los casos, las situaciones no se caracterizan por ser ideales. Podemos encontrarnos, por ejemplo, con que toda la información esencial no pueda ser reflejada de forma breve; aunque lo que realmente representa un problema son las distintas percepciones que tienen diferentes personas de los conceptos *breve* o *esencial*. Esta concepción vendrá marcada por los diferentes intereses de cada uno o la utilidad que se le vaya a dar al resumen obtenido. Lo que para un individuo o grupo puede resultar interesante o esencial, puede no serlo para otros, y viceversa, incluso se pueden presentar visiones diferenciadas acerca de qué extensión puede ser considerada o no como breve.

El proceso de resumir es una actividad inherente al ser humano. Las personas continuamente reciben información del exterior que someten a una serie de procesos (transformación, reducción, almacenaje, recuperación o utilización) según sus capacidades y necesidades, con vistas a una futura aplicación. Podemos considerar pues que resumir implica una actividad de reducción natural de información en la mente humana. En este proceso se pasa a fijar los conceptos más importantes o significativos de entre todos los datos percibidos. En general el resumen pretende ser lo mismo pero en tamaño más pequeño y no una parte arbitraria de lo que tenemos que resumir.

La obtención de un buen resumen es algo muy importante. No basta con tener un resumen, dicho resumen debe de satisfacer las necesidades del usuario y debe de hacerlo con una cierta calidad. La selección de características es una acción que determinará qué partes son las que queremos destacar o a cuáles prestar mayor atención. Una vez hecho esto y obtenido el resumen final, se deberá de medir la calidad de resumen, a fin de que podamos conocer cómo de bueno es el producto obtenido.

La tarea de intentar que las máquinas produzcan resúmenes comparables a los realizados por humanos, es una tarea de gran interés en la sociedad actual. El punto fuerte de los ordenadores en esta tarea lo aportan su capacidad para almacenar datos y su potencia de cálculo que puede realizar millones de operaciones en un tiempo muy inferior al que nosotros emplearíamos. En cambio, el punto débil es el de conseguir que la información que produce sea comprensible para los humanos, que en la mayoría de las ocasiones son los receptores habituales.

Pero algo que para los seres humanos es tan natural y que incluso se realiza sin reparar en ello, no es una tarea tan fácil como puede parecer. Los humanos adaptamos inconscientemente nuestra forma de actuar y desenvolvemos según en entorno y la situación en la que nos encontremos. Cuando nos enfrentamos al reto de intentar que las máquinas realicen resúmenes comparables a aquellos que producimos nosotros mismos, nos damos cuenta de ello.

Debido a la dificultad y utilidad del tema un gran número de investigadores está centrando sus estudios en mejorar el proceso de comunicación entre humanos y máquinas. En nuestro caso, son de especial interés los esfuerzos de investigación realizados en el ámbito del resumen de series de datos temporales.

III. LA IMPORTANCIA DE LA DIMENSIÓN TEMPORAL

Desde los tiempos más remotos el ser humano ha medido el paso del tiempo con diferentes métodos y herramientas, algunos de ellos de gran precisión. A pesar de ello, el estudio de las series de tiempo en sí, posee un origen relativamente reciente. Se piensa que fue hace aproximadamente 1000 años cuando se produjo la primera representación gráfica de los eventos dividiendo un eje horizontal en intervalos de igual amplitud para representar iguales periodos de tiempo.

En este trabajo seguiremos las ideas de Peña cuando afirma que “una serie temporal es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares (cada día, cada mes, cada año, etc.)” [24]. Las primeras series temporales estudiadas correspondían a datos astronómicos y meteorológicos.

Las series de datos temporales pueden ser representadas mediante una sucesión de las medidas tomadas. Si estas medidas no se han tomado en intervalos regulares o necesitamos obtener más información acerca del momento de tiempo al que corresponde, dichas medidas pueden ser acompañadas por el instante de tiempo concreto con mayor o menor nivel de detalle dependiendo de nuestras necesidades. Sin embargo, en ocasiones estas formas de representación, bien sea en texto plano o ayudados mediante tablas, no suelen ser muy intuitivas.

En algunas ocasiones puede que el usuario que recibe la información no posea conocimiento experto en el tema específico. Otras veces, puede que la cantidad de datos sea tan elevada o la diferencia entre sus valores tan notable, que hagan complicado el proceso de análisis de los datos. Sea como fuere, incluso con las series de datos más sencillas, en muchos de los casos la representación gráfica aporta una buena herramienta de representación de las series de datos temporales. Por desgracia, la representación gráfica de las series no siempre es fácil de interpretar, ya que en ocasiones las series son muy complicadas o incluso tenemos varias series relacionadas entre sí presentadas en el mismo gráfico. A esto se suman problemas como la necesidad de equipos apropiados y las diferentes percepciones dependiendo de la granularidad usada.

En todos los casos, con independencia de la complejidad de la serie o series, el resumen lingüístico de series de datos temporales es un herramienta potente que permite presentar a usuarios no expertos información acerca de la serie en un formato comprensible y fácil de interpretar. Otra buena característica atribuible a los resúmenes lingüísticos es la posibilidad de uso de un sintetizador de voz, de especial utilidad en aquellos casos en los que la visualización de las gráficas no es adecuada o no es posible (por ejemplo, para poder presentarse los resultados a personas con algún tipo de problema visual).

La importancia del tiempo es crucial cuando se analizan datos en las empresas, por ejemplo, para asesorarse en el proceso de toma de decisiones. De hecho, el tiempo juega un papel muy importante en los almacenes de datos, o data warehouses, debido a que esta dimensión permite almacenar la información histórica concerniente a las operaciones diarias

de una organización. De este modo, el personal encargado de la toma de decisiones puede realizar el análisis de la evolución de diferentes y variados aspectos del negocio a lo largo del tiempo. Por ello es por lo que la dimensión temporal suele estar presente casi siempre entre las dimensiones de los cubos de datos.

De forma breve y general se puede describir un data cubo o cubo de datos como un conjunto multidimensional de celdas. Cada una de las dimensiones que conforman dicho cubo pueden ser vistas como un conjunto de miembros que pueden organizarse a su vez haciendo uso de una o varias jerarquías. Gracias a la organización jerárquica de las dimensiones, la granularidad en el cubo de datos puede variar en cada dimensión dependiendo del nivel seleccionado en cada momento. Las operaciones OLAP permiten a los usuarios aprovechar la ventaja de esta particularidad y consultar de diferentes formas al data cubo con el fin de obtener los datos deseados en cada momento.

Podemos obtener series de datos temporales mediante observación directa pero es más usual hacerlo mediante recogida automática de datos con volcado de la información en sistemas de almacenamiento. Si hacemos consultas sobre este tipo de estructuras de almacenamiento seremos capaces de obtener series de datos temporales. Si además las estructuras son multidimensionales como las mencionadas anteriormente, mediante la aplicación de operaciones OLAP seremos capaces de obtener series temporales con diferentes granularidades.

La importancia de la dimensión temporal y la necesidad de obtener resultados entendibles por los usuarios humanos, hacen de este tipo de series candidatas principales para beneficiarse de técnicas de resumen lingüístico en su estudio.

El propósito del estudio o análisis de las series de datos temporales puede ser dividido en dos grandes áreas según varios autores. La primera de ellas es “entender o modelar el mecanismo estocástico de una serie observada” y la otra “predecir o pronosticar los valores futuros de series basadas en la historia de esas series y, posiblemente, otras series o factores relacionados” (Cryer [11]).

De este modo, podemos decir que el análisis de series temporales comprende métodos que ayudan a interpretar los datos, extrayendo para ello información representativa, tanto referente a los orígenes o relaciones subyacentes como a la posibilidad de extrapolar y predecir su comportamiento futuro. De hecho uno de los usos más habituales de las series de datos temporales es su análisis para predicción y pronóstico. En nuestro caso, no vamos a centrar nuestra atención en áreas como la predicción o pronóstico, sino en la posibilidad de la descripción de las series.

La tarea de obtención de la información subyacente en las series de datos temporales es importante y permite conocer tendencias, eventos destacados o patrones, que permitirán una mejor toma de decisiones. Si esta nueva información se encuentra representada en formato textual con patrones similares a los usados por los humanos cuando se comunican entre ellos, la toma de decisiones se desarrollará de forma más amigable para el decisor.

IV. RESUMEN LINGÜÍSTICO DE DATOS CON TÉCNICAS DIFUSAS

La lógica difusa tiende un puente entre los datos tal y como están almacenados en las máquinas, y las descripciones lingüísticas con las que los humanos están acostumbrados a tratar. La comunicación humano - máquina es uno de los temas más candentes en la actualidad. Los resúmenes lingüísticos se sirven de las bondades de la lógica difusa para conseguir un proceso de comunicación más eficaz.

En trabajos como [20] de Mitra et al. y [9] de Chen et al. podemos ver completos estados del arte de técnicas de obtención de información a partir de grandes cantidades de datos mediante el uso de técnicas Soft Computing.

Dado el interés que posee la realización de este tipo de resúmenes existen distintos enfoques a la hora de afrontar la tarea de producir un resumen lingüístico de datos. Como denominador común cabe destacar que la gran mayoría de las técnicas hasta ahora desarrolladas hacen uso de etiquetas lingüísticas. Las etiquetas lingüísticas son esenciales ya que introducen la posibilidad de tratar con la imprecisión y vaguedad necesarias para trabajar con el lenguaje natural.

Son varios los autores que se han dedicado a trabajar en este campo. Ronald R. Yager lo aborda apoyándose en el uso de sentencias enriquecidas por conceptos difusos [25] introducidos años antes por Lofti A. Zadeh. Estas sentencias cuantificadas representan el grado satisfacción de una cierta cualidad por un grupo de objetos. En un primer momento, la cuantificación de las sentencias se realizaba usando el cardinal de Zadeh, pero posteriormente, y para la resolución de casos más complejos, desarrolló los denominados operadores OWA (Ordered Weighted Averaging) [26] y [27].

La investigadora Anne Laurent presentó sus técnicas en [19]. En su trabajo, se centra en la obtención de resúmenes lingüísticos a partir de bases de datos multidimensionales y operaciones OLAP, usando para ello una interfaz amigable para el usuario. En [29] de Zhang los resúmenes lingüísticos se realizan usando Degree Theory y FCA (Formal Concept Analysis). El autor diferencia entre el proceso a seguir para obtener lo que él denomina resúmenes simples y resúmenes complejos.

Anteriormente ya ha sido mencionado el fuerte vínculo que muchos autores comienzan a ver entre las técnicas de minería de datos y la obtención de resúmenes lingüísticos. Un trabajo más profundizando en esta reflexión lo encontramos en [17] de Janusz Kacprzyk and Slawomir Zadrozny.

Enfatizando la importancia de los resúmenes lingüísticos, pero sobre todo los resúmenes lingüísticos de series de datos temporales encontramos el trabajo de Ichiro Kobayashi y Naoko Okumura [18]. En el mencionado trabajo los autores muestran su interés en la verbalización de series de datos temporales mediante técnicas interesantes. En [10] Chiang et al. también expresan su interés en el resumen lingüístico de series de datos temporales.

El concepto de resumen lingüístico es también abordado en los trabajos de Díaz et al. [14], Triviño et al. [15] y Moreno [16].

V. HACIA UNA VERSIÓN LINGÜÍSTICA DE FCUBE FACTORY

Durante esta sección se hará una pequeña introducción a los fundamentos de nuestro enfoque para el resumen lingüístico de series de datos temporales. Una vez realizado el repaso de las ideas más importantes, se presenta la introducción de nuestra herramienta en una plataforma web que permite el manejo de cubos de datos.

Nuestras contribuciones en el área pueden ser consultadas en [4], [5], [7], [6], [8], [3]. En nuestros trabajos tomamos como punto de partida una serie de datos temporales obtenida a través de la aplicación de operaciones OLAP sobre un cubo de datos multidimensional. Como consecuencia encontramos que la dimensión tiempo se encuentra descrita con diferentes niveles de granularidad, dando lugar esto a una partición jerárquica difusa. Además de la nombrada partición, tenemos otra partición difusa en la dimensión del hecho que queremos describir. Dichas particiones cumplen una serie de características (consultar artículos para más información). Los resúmenes finales se componen de sentencias cuantificadas de tipo II, *Q de los D son A* donde Q es un cuantificador, D y A son conjuntos difusos representando el tiempo y el hecho. Un ejemplo: *La mayoría (Q) de los días en la estación cálida (D) las precipitaciones son escasas (A)*.

Vamos a verlo más claro un con ejemplo. La Figura 1 muestra una serie de datos temporal y su contexto lingüístico. La serie representa la afluencia de pacientes a un centro de salud dado durante un periodo de tiempo igual a un año. Como se ha comentado anteriormente las dos dimensiones, tanto la de tiempo como la de la variable a estudiar, están particionadas usando lógica difusa. En este caso, las particiones pueden haber sido heredadas de un data cubo multidimensional o simplemente facilitadas por el usuario que desea obtener el resumen. La dimensión tiempo está definida mediante una jerarquía de particiones difusas. Cada una de las particiones con diferente granularidad, o lo que es lo mismo, con diferente grado de abstracción.

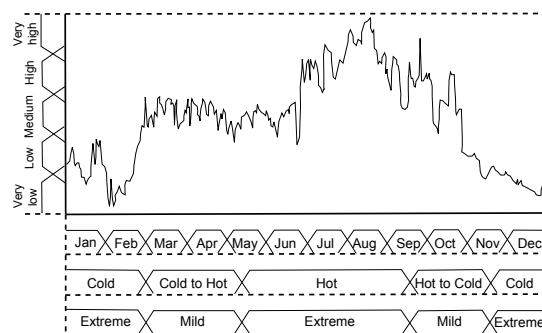


Fig. 1. Ejemplo de serie de dato temporal y su contexto lingüístico.

A partir de una serie como la de la Figura 1 y considerando el contexto lingüístico, así como otros parámetros de configuración, nuestro método sería capaz de obtener resúmenes como el mostrado a continuación:

Most of the patient inflow is low or very low with cold weather; medium in seasons with cold to hot

weather, May, and June; very high or high in July and August; and high or medium in September and October. Finally, at least 70% of patient inflow is low or very low in November.

FCube Factory es una plataforma web que nos permite la manipulación de grandes cantidades de datos almacenados siguiendo el modelo multidimensional ([12], [22]). En su versión inicial se implementaron diversas técnicas que ofrecían información novedosa obtenida a partir de los datos almacenados. Las capacidades de esta potente plataforma se están ampliando en la actualidad mediante la adición de la posibilidad de obtención de resúmenes de datos en general y de series temporales en particular.

A través de la plataforma se pretende que el usuario tenga la posibilidad de consultar cubos de datos multidimensionales de forma rápida y sencilla, presentando los datos con diferentes formatos (incluyendo la síntesis de audio a partir de los resúmenes), pudiendo consultar el conjunto de sentencias cuantificadas que conforman el resumen o un texto completo fruto del postprocesado de dichas sentencias.

El análisis de la variación de afluencia, esta vez a otro centro de salud, sigue siendo el tema en el siguiente ejemplo. En esta ocasión los resultados se han tomado directamente de la herramienta FCube Factory (ver Figura 2).

La pantalla está dividida en tres grandes zonas de presentación de resultados. En primer lugar, se muestra al usuario una representación gráfica de la serie de datos temporal. A continuación, se presenta el resumen de texto postprocesado con la posibilidad de la reproducción del audio asociado. En tercer lugar se ofrece la posibilidad de analizar cada una de las sentencias que componen el resumen final. Al seleccionar una de las sentencias marcarán las zonas afectadas en el gráfico y se ofrecerá la posibilidad de reproducir el audio de la sentencia individual. Destacar que la posibilidad de asociar cada una de las sentencias a los datos que soportan la afirmación en el gráfico, ofrece al usuario una herramienta para ponerse en contexto y comprobar que la veracidad de la afirmación.

La posibilidad de contar con una herramienta visual y auditiva, sencilla de manejar que nos presente los resultados deseados siguiendo los patrones de lenguaje usado por los humanos, es un apoyo muy grande sobre todo en áreas tan importantes como el soporte a la decisión.

VI. CONCLUSION

En la actualidad es muy necesario ser capaz de manejar grandes conjuntos de datos. No es extraordinario el hecho de tener que enfrentarse a una situación como la descrita, si no frecuentemente, sí alguna vez en la vida. Las empresas, organizaciones e instituciones dependen en gran medida de llevar este proceso a cabo de forma eficaz.

Es vital el manejo de importantes volúmenes de datos, pero más lo es el hecho de ser capaz de extraer información útil de dichos datos. Una de las dimensiones más reflejadas en estos datos es la dimensión temporal. Las empresas mantienen datos de su producción, las ventas, los gastos, el stock de productos, y tantas otras facetas similares, que reflejan su funcionamiento y su evolución a lo largo del tiempo.

Se ha identificado la necesidad de introducir el resumen lingüístico en este campo con el fin de conseguir métodos rápidos, sencillos y que produzcan resultados informativos, de calidad y amigables para los usuarios. Los resultados obtenidos siguiendo patrones semejantes a los del lenguaje natural serán mucho más legibles y por lo tanto útiles para su posterior uso.

Los resúmenes lingüísticos de series de datos se presentan como una herramienta de gran utilidad en el ámbito del Business Intelligence en general y en el apoyo a la toma de decisiones en particular.

REFERENCES

- [1] I. Z. Battyshin and L. Sheremetov. Perception-based approach to time series data mining. *Appl. Soft Comput.*, 8(3):1211–1221, 2008.
- [2] I. Z. Battyshin and T. Sudkamp. Perception based data mining and decision support systems. *Int. J. Approx. Reasoning*, 48(1):1–3, 2008.
- [3] R. Castillo-Ortega, N. Marín, and D. Sánchez. Linguistic query answering on data cubes with time dimension. *Special Topic Issue on Advances in Fuzzy Querying: Theory and Applications in International Journal of Intelligent Systems (IJIS)*, page To appear.
- [4] R. Castillo-Ortega, N. Marín, and D. Sánchez. Fuzzy quantification-based linguistic summaries in data cubes with hierarchical fuzzy partition of time dimension. In H. Yin and E. Corchado, editors, *IDEAL'09*, volume 5788 of *LNCS*, pages 578–585. Springer, Heidelberg, 2009.
- [5] R. Castillo-Ortega, N. Marín, and D. Sánchez. Linguistic summary-based query answering on data cubes with time dimension. In T. Andreasen, R. R. Yager, H. Bulskov, H. Christiansen, and H. L. Larsen, editors, *FQAS'09*, volume 5822 of *LNAI*, pages 560–571. Springer, Heidelberg, 2009.
- [6] R. Castillo-Ortega, N. Marín, and D. Sánchez. Linguistic local change comparison of time series. In *Congreso Español de Informática. CEDI 2010*, pages 451–459, 2010.
- [7] R. Castillo-Ortega, N. Marín, and D. Sánchez. Time series comparison using linguistic fuzzy techniques. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2010*, pages 330–339, 2010.
- [8] R. Castillo-Ortega, N. Marín, and D. Sánchez. A fuzzy approach to the linguistic summarization of time series. *Special Topic Issue on Soft Computing Techniques in Data Mining in Journal of Multiple-Valued Logic and Soft Computing (JMVLS)*, 17(2,3):157–182, 2011.
- [9] G. Chen, Q. Wei, and E. E. Kerre. *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques, Massive Computing Series*, chapter 14 Fuzzy Logic in Discovering Association Rules: An Overview, pages 459–493. Massive Computing Series. Springer, Heidelberg, Germany, 2006.
- [10] D. Chiang, L. R. Chow, and Y. Wang. Mining time series data by a fuzzy linguistic summary system. *Fuzzy Sets Syst.*, 112:419–432, June 2000.
- [11] Jonathan D. Cryer and Kung-Sik Chan. *Time Series Analysis with applications in R*. Springer, 2008.
- [12] M. Delgado, C. Molina, L. Rodríguez Ariza, D. Sánchez, and M. A. Vila Miranda. F-cube factory: a fuzzy olap system for supporting imprecision. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(Supplement-1):59–81, 2007.
- [13] María Pinto Molina et al. *Aprendiendo a resumir. Prontuario y resolución de casos*. TREA, Gijón, 1 edition, 2005.
- [14] A. Bugarín, F. Díaz-Hermida. Linguistic summarization of data with probabilistic fuzzy quantifiers. In *ESTYLF 2010*, pages 255–260, 2010.
- [15] A. S. Montemayor J. J. Pantrigo R. Cabido G. Triviño, A. Sánchez and E. G. Pardo. Linguistic description of traffic in a roundabout. In *IEEE International Conference on Fuzzy Systems - FUZZ-IEEE (WCCI 2010 IEEE World Congress on Computational Intelligence)*, pages 2158–2165, 2009.
- [16] J. Moreno García, J. J. Castro-Schez, and L. Jiménez. A fuzzy inductive algorithm for modeling dynamical systems in a comprehensible way. *IEEE T. Fuzzy Systems*, 15(4):652–672, 2007.
- [17] J. Kacprzyk and A. Wilbik. Using fuzzy linguistic summaries for the comparison of time series: an application to the analysis of investment fund quotations. In U. Kaymak J. P. Carvalho, D. Dubois and J. M. C. Sousa, editors, *IFSA-EUSFLAT 2009*, pages 1321–1326, 2009.
- [18] I. Kobayashi and N. Okumura. Verbalizing time-series data: With an example of stock price trends. In *IFSA/EUSFLAT Conf.*, pages 234–239, 2009.

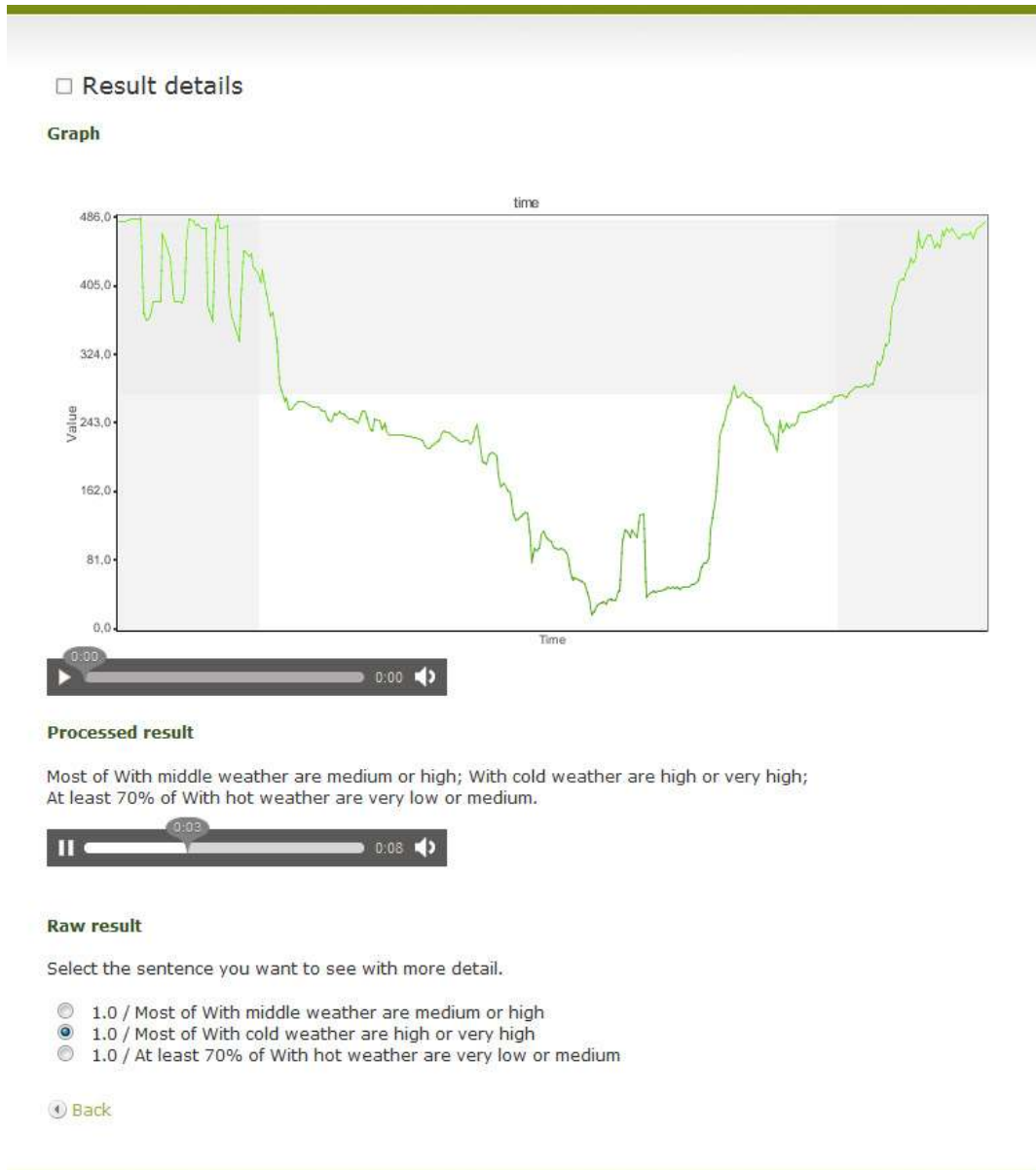


Fig. 2. Captura de pantalla de un resumen lingüístico en FCube Factory.

- [19] A. Laurent. A new approach for the generation of fuzzy summaries based on fuzzy multidimensional databases. *Intell. Data Anal.*, 7:155–177, April 2003.
- [20] S. Mitra, Senior Member, Fellow, S. K. Pal, and P. Mitra. Data mining in soft computing framework: A survey. *IEEE Transactions on Neural Networks*, 13:3–14, 2001.
- [21] D. Mladenic, N. Lavrac, M. Bohanec, and S Moyle, editors. *Data Mining and Decision Support*, volume 745 of *The Springer International Series in Engineering and Computer Science*. 2003.
- [22] C. Molina, L. Rodríguez Ariza, D. Sánchez, and M. A. Vila Miranda. A new fuzzy multidimensional model. *IEEE T. Fuzzy Systems*, 14(6):897–912, 2006.
- [23] J. A. O’Brien and G. M. Marakas. *Management information systems*. McGraw-Hill, 8 edition, 2008.
- [24] Daniel Peña. *Análisis de Series Temporales*. Alianza, 2005.
- [25] R. R. Yager. A new approach to the summarization of data. *Information Sciences*, (28):69–86, 1982.
- [26] R. R. Yager. Toward a language for specifying summarizing statistics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 33(2):177–187, 2003.
- [27] R. R. Yager. A human directed approach for data summarization. In *IEEE International Conference on Fuzzy Systems*, pages 707–712, 2006.
- [28] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8, 1965.
- [29] L. Zhang, Z. Pei, and H. Chen. Extracting fuzzy linguistic summaries based on including degree theory and fca. In *Proceedings of the 12th international Fuzzy Systems Association world congress on Foundations of Fuzzy Logic and Soft Computing*, IFSA ’07, pages 273–283, Berlin, Heidelberg, 2007. Springer-Verlag.

*fuzzy*BSC for Quality Assurance Learning Process

Luis Cerda*, Juan Horacio Campos[†], Mario Rojas[†], and Daniel Sánchez[‡]

*Facultad de Ciencias Empresariales, Universidad Autónoma de Chile,
e-mail: luis.cerda@uautonoma.cl

[†]Facultad de Economía y Negocios, Universidad Andrés Bello Chile,
e-mail: juanh.campos@uandresbello.edu,

[†]Facultad de Economía y Negocios, Universidad Andrés Bello Chile,
e-mail: jm.rojas@unab.cl.

[‡]European Centre for Soft Computing, Mieres, Asturias, Spain, and
with the Dept. Computer Science and A.I., University of Granada, Spain.
e-mail: daniel.sanchezf@softcomputing.es.

Resumen: El Fuzzy Balanced Scorecard ha sido propuesto por Bobillo, Delgado, Gómez-Romero y López [BF09], para gestionar ámbitos caracterizados por la imprecisión y la ambigüedad, en este trabajo se muestra su aplicación al aseguramiento de la calidad del proceso formativo universitario de pre-grado y lo complementa con la extracción de reglas de asociación difusa para mantener la vigencia de los indicadores gestión y sus metas.

Palabras claves: Reglas de Asociación, Minería de datos, Gestión y Aseguramiento de la Calidad, Proceso de Enseñanza-Aprendizaje, Lógica Difusa.

Abstract— The Fuzzy Balanced Scorecard has been proposed by Bobillo, Delgado, Gomez-Romero and Lopez, to manage areas characterized by the vagueness and ambiguity, this paper demonstrates its application to ensure quality of university learning process of undergraduate and supplements it with the extraction of fuzzy association rules to maintain the validity of the management indicators and its goals.

I. INTRODUCTION

THE high and increasing competitiveness of the Chilean educational system in the third cycle requires its suppliers to demonstrate that the learning process made is with quality certified. Ensuring the quality of the learning process involves handling the vagueness and ambiguity inherent in the evaluation and monitoring of performance and efficiency of its protagonists (students and teachers), which represents a challenge not overcome by conventional management tools.

Among the management tools to assess performance, it is recognized the Balanced Scorecard proposed by Norton and Kaplan [1], as one of the more efficient pro-active management tools in the business sector. It is also known

that the model in its original version is not apt to manage environments where the performance measurement introduces fuzzy variables for this purpose Bobillo et. al. [2] proposed the so-called fuzzy Balanced Scorecard, from now on, fuzzyBSC.

This paper shows that this emerging perspective fuzzyBSC, is valid as an alternative of solution for the continuous and systematic control of the learning process, given that it handles vagueness and uncertainty and also provides proactive support for decision-making leading to ensure quality.

The fuzzyBSC model is reinforced extracting useful and novel knowledge from historical databases that store transactions that the learning process generates, which permits to verify the validity of the processing rules, rates and targets.

The fuzzyBSC is updated applying data mining techniques, it permits to control more accurately the level of progress and degree of development of cognitive, behavioral, attitudinal and procedural skills that reaching students during the learning process, the level and quality of faculty contribution; level of completeness of the

profile of egress and their relevance to the needs of the environment and of potential employers, finally the level of satisfaction with the learning experience gained and the grade of fulfillment of expectations both in terms employability and in the level of monthly income received by the graduate.

Thus, the solution *fuzzyBSC* for Quality Assurance Learning Process, permits to solve two of the main problems that represents pro-active management of the learning process, namely, on one hand having a managerial tool that allows to objectify the subjectivity to which are measured the entry skills, the efficiency of the process of skills development, verification of the result obtained in each of the stages leading to creation of the egress profile and the stage of success obtained by the new professional when he incorporates himself into the milieu and the other hand, to maintain current strategic objectives, its management indicators and its goals, extracting useful and novel knowledge from the historical database that provides the same solution.

This document presents the preliminary work for free adaptation of *fuzzyBSC* to quality assurance of the learning process and the verification of validity of the processing rules through data mining techniques and it is established the future work to develop to implement it as solution for the quality assurance of the learning process of an academic unit of undergraduate.

This paper is organized as follows: Section 2 shows the methodologies and main conceptual aspects of this work, Section 3 describes the current status of field of study, Section 4 presents an example, formulating a dimension of *fuzzyBSC*, the Section 5 shows the verification process of the indicators and *fuzzyBSC* goals through extraction of mining fuzzy association rules, finally Section 6 contains some conclusions of the work done and some ideas for future work.

II. BACKGROUND

A. Quality Indicators of the Learning Process

According to a study made by Fernández et. al. [3], international guidelines used to determine the quality of higher education institutions, having considered the overall coverage of the ranking published by The Times Higher University Ranking; Academy Ranking of World Universities; World Universities Ranking on the web, Asia's Best Universities 2000, The Center Research Universities Ranking and The Sunday Times University Guide, notice that, in general, the most used indicators are:

- o Teachers/students ratio.
- o Entry Qualifications, ie, high school grades and

university selection tests.

- o Number of citations and/or publications (adjusted for the number of equivalent full-time teachers)
- o Peer evaluation and/or evaluation of academic excellence.

While in the local context, an academic unit-level or professional career and in order to guide applicants to tertiary education, the state through the Ministry of Education, offers the student community and the wider public, trends and patterns that show the university careers as the following indicators:

- o Enrolment number and last score entered
- o Level of retention
- o Actual duration of the career
- o Level of employability of graduates
- o Monthly Income from the first and fifth year out of University

In this paper we include data from the main reference point, the report "Working Future 2009/2010" sponsored by the Ministry of Education, University of Chile and Diego Portales University, and coordinated by the eminent national experts Patricio Meller and José Joaquín Brunner.

B. Learning Process Quality Control

Although the indicators identified in the previous section, of one way or another way, sanction the quality of the result of the learning process, they do not control it, for these effects and in this sense, for this work is considered, as benchmark, the systematic model of Education System Quality, developed by Arturo De la Orden [4], which establishes the relationship of effectiveness, efficiency, functionality, level of use and allocation levels between inputs, functions, goals and objectives and education products. It raises the consistency between inputs, processes, products, goals and satisfaction of social needs and expectations that define the quality of education as functionality, relevance.

Secondly, it notes the consistency of the product with the goals and objectives, defines the quality of education as efficiency or effectiveness.

Third, the consistency between, on one hand, input and process, and other hand, products, defines the quality of education as efficiency.

As a summary indicates that the explanatory concept of the assessment of the quality of education jointly considers the three interrelated dimensions: functionality, efficiency and effectiveness.

C. Step by Step Evaluation of the Learning Process

In Arturo de la Orden's aforementioned work, it is noted that the Joint Committee on Standards for Educational Evaluation, believes that "the evaluation is the systematic determination of the value or merit of an object", the concept "value" clearly indicates that any evaluation involves a value judgment and therefore it may not be pretended neutrality, on the contrary, requires a defensible set of principles or ideals to determine the status of the object of being evaluated with respect to such values. Often the value of object to be evaluated is specified on its merit or intrinsic value, this accounts for the subjectivity and vagueness inherent in the evaluation process that this paper argues.

It also notes that the intended use of the results of the evaluation (information and value judgment of the object), distinguishes two categories called learning function and summative function, where the first is an evaluation of processes and the summative function is a final evaluation. To formative evaluation is given the proactive nature and whose function could be identified with permanent decisions-making on program processes, thus considering the summative evaluation as retro-active with the function of demanding responsibility or accountability as a consequent of the final results of the program, which tends to confirm the proposal to have a mechanism to permit monitoring step by step the learning process.

With regard to the concept of quality, De la Orden states "The lack of a concept of quality as a benchmark for evaluation of an educational system obviously is translated into the impossibility of establishing criteria for the selection of variables as indicators, ensuring information about something we call quality, whose identity is not determined", which leads us to propose the verification of indicators and targets by applying data mining to verify their relevance

He also notes that the connection with the movement focused on the management and quality control and total quality, developed and promoted by the business world, a concept that in the educational field has a high level of ambiguity and scant operating as educational activity guide, which is remedied with the concepts introduced by Bobillo in his *fuzzyBSC* model.

Achieving the egress profile with effective duration of the similar career to the theoretical time required to control step by step the process, for this measurement and quality assurance of the learning process, the authentic evaluation is the emergent evaluation process that leads to guarantee it.

D. Authentic Evaluation of the Learning Process

The demands of the environment requires that vocational learning is comprehensive, meaning that learning and its evaluation also covers the cognitive aspect of yesteryear, whose evaluation was binary long time (Knows/Does not know), covering behavioral, attitudinal and procedural fields of the professional profile, to that end, the emerging technique of evaluation in faculty who have accepted the challenge, is the authentic evaluation, which is aimed at assessing the level of skill development through action manifested, which, it claims to do these measurements with wider tasks such as troubleshooting. In assessing the domain of certain skills and competence should be applied an expert opinion, which as in all areas, has a strong element of intuition and subjectivity of the evaluator.

E. Intuition, Subjectivity, and Vagueness of the Learning Process Evaluation

As Arturo de la Orden's work aforementioned points out, evaluation of the result of each stage takes the form of merit or intrinsic value, which accounts for the intuition, subjectivity, vagueness and ambiguity that are inherent to the learning evaluation process, the analysis of the vagueness and ambiguity dates back to early last century and is one of its highest expressions in 1965 with the work of Lotfi Zadeh [5], with the definition of fuzzy set that gives rise to the Theory of Possibility. This theory allows expressing in mathematical terms the approximate reasoning which is used by man in everyday life, describing facts that are not totally true or totally false, for that it is not apply the classical logic or logic bivariate. This theory of Fuzzy Logic allows using concepts concerning to reality, defining varying grades of membership and following reasoning patterns similar to those of human thought, Kosko [6]. For example, in the formal logic if a result is good, it can be score "achieved", and otherwise "not achieved", namely, true or false, approved or disapproved, for fuzzy logic is possible to identify intermediate outcomes such as competition grades as proposed by Gil Aluja [7] and interpretation of grades that notes the regulation of the student, both are reflected in Table I.

In this sense, as manifested by Bobillo, the membership functions used for the definitions of fuzzy sets such as triangular and trapezoidal functions are fully applicable in this context.

Thus, from the standpoint of the learning process, this model permits to capture more accurately the reality in terms of reflecting in best way the vagueness and subjectivity of the evaluations.

TABLE I
INTERPRETATION OF EVALUATION

Approved	Levels of competition	Grades
Yes	1.0 Fully competent	7 Excellent 6 Very Good 5 Good 4 Sufficient
	0.9 Competent	
	0.8 Practically competent	
	0.7 Enough competent	
	0.6 Nearly Competent	
	0.5 Moderately competent	
No	0.4 Almost incompetent	3. Insufficient 2. Deficient 1. Bad
	0.3 Quite incompetent.	
	0.2 Practically incompetent	
	0.1 Incompetent	
	0.0 Totally incompetent	

F. The Relationship Between Authentic Evaluation and Fuzzy Logic.

The authentic evaluation requires to objectify the subjectivity associated with measuring the performance of individuals participating in the learning process, this objectification is achieved by fuzzy logic, which converts input variables (qualitative and quantitative) on linguistic variables through membership functions or fuzzy sets. A linguistic variable is a variable whose values are words or sentences (not numbers). Often you want to describe the state of an object or phenomenon, for that, it is used a variable whose value does the description, for example in our case, to assess the level of proficiency achieved, this is measured in terms of ability, disposition or willingness to perform it and the knowledge applied when the particular task is performed, this is solved using fuzzy variables that admits their values are linguistic labels whose terms are defined as fuzzy sets (on some underlying domain).

The incorporation of linguistic labels, which, by its affinity with the way of human's reasoning, have characteristics appropriate to meet this challenge in addition to granting the ease of storing and processing historical data to extract further useful and novel knowledge hidden by fuzzy association rules which they provide the best result in similar situations as described by Delgado et. al. [8].

G. Maintaining the fuzzyBSC through Data Mining and Fuzzy Association Rules

Data mining, aimed at obtaining useful knowledge, not explicit from the data stored in repositories with historical data, represents a methodology that offers an excellent alternative of solution to the problem under study, since, as in other areas, is recognized that the extraction of

knowledge is a basic need, in our specific case, to determine trends and patterns in a field in constant evolution and improvement that nowadays requires timely and effective decisions.

The treatment, storage and processing of historical data with their respective linguistic labels for each results obtained by its protagonists, in different learning activities, are contributing the attributes that make a transaction in the repository of historical data.

The evaluation of each stage of the learning process, the evaluated task characteristics, the context of evaluation and other specific materials to the evaluator and who is evaluated, all of them form a set of items, given this set of items, I, and a set of transactions T (which is also sometimes called T-set), where each transaction is a subset of I, we say that an association rule [9] is an "implication" of the form $A \rightarrow C$ that denotes the presence of the itemsets (or setof sets) A and C in certain transactions of T, assuming that $A, C \subset I$, $A \cap C = \emptyset$; and $A, C \neq \emptyset$. We shall follow the notion of fuzzy transactions and fuzzy association rules introduced in [8], employing support and certainty factors for assessing the rules. We shall also employ the approach of very strong rules proposed in [10], that states that a fuzzy association rule is very strong if both the rule and its contrapositive are strong.

H. BSC as a Generalized Model of Management Control

The control of management or performance management nowadays has an implicit relationship with the instrument presented in 1992 at the Harvard Business Review named the Balanced Scorecard by Norton and Kaplan [1]. The objective of this instrument is continuously evaluating the organizational, group and individual performance, from four perspectives:

1) *The perspective of organizational dynamics* which permits to evaluate the performance of members of the organization and its commitment to the vision, mission and strategic objectives of the organization and this performance is the basis for the following perspectives.

2) *The perspective of internal processes* which permits to determine whether the performance of internal processes serves to influence customers.

3) *The perspective of customers* which allow knowing if the target segment is getting the added value it is expected to receive.

4) *The financial perspective* (which in our case

corresponds to the perspective of user/beneficiary when dealing with non-profit entities as stated by María José González Quintana and Encarnación Cañadas in [11]), which permits to determine the added value to stakeholders and the positive externalities that generates the environment and society in general.

I. BSC as a Management Control Model of the Learning Process

In the work published by Leonel Arias et. al. [12], it is raised the application of the BSC in public higher education institutions and they build the strategic map considering for Human Capital perspective and Organizational Dynamics evaluate the alliance with other institutions, the learning programs, technology and know-how (Know How), the internal process perspective, consider the research, extension, administrative modernization and academic reform among others, on the customer perspective where applicants, students, graduates and employers are included, consider efficiency and effectiveness in research, extension and teaching and finally in the Financial perspective or Contribution of Interest Groups consider the community, public and private companies, and other indicators such as the egress of skilled professionals, knowledge transfer to society and improving quality of life of the community, in our case we consider to graduates and potential employers are part of the users/beneficiaries scale.

J. Fuzzy BSC Applied to the Learning Process

Bobillo's work et. al., to incorporate the treatment of uncertainty and ambiguity to the BSC, transforms it and presents it as a valid option for continuous evaluation of the performance of the different roles and different stages of the learning process that this work aims to obtain.

The proposed design for the different perspectives of fuzzyBSC to continuously evaluate the learning process under study is the following:

1) The basis perspective is of organizational dynamics (Faculty and Staff): The strategic objective of the teaching unit is to have excellence faculty and staff. Indicators that permit to verify fulfillment with these objective, are, for example teachers/students ratio, the number of citations and/or publication of full-time teachers and those that measure the organizational climate conducive to excellent learning experience of their students.

The organizational climate is measured in terms of attendance and punctuality of teachers. Teacher's

contribution is measured in terms of the progress of implementation of commitments by teachers and the result of the evaluation by the students of teachers' work in relation to the validity and appropriateness of content and activities developed by teachers in their subject, among others.

2) For the perspective of internal processes providing an efficient and effective learning process with good indicators that allow the level of progress in content and activity of the subject that lead to achievement of the professional profile of the student and to determine the level of retention timely (attendance of pupils) and risk conditions of repeat that attacks against the actual duration of the career, this requires evaluating the development attained in skills and knowledge by students and by fulfillment of progress level of the program of subject, timeliness and completeness of content delivery and implementation of planned activities and evaluations, among others.

3) For the perspective of customers (students) providing excellent learning to attract the best applicants and students getting good indicators in number of stakeholders and last registered scores and good grades from the student and teacher, that is measured in terms of satisfaction for the learning experience of students and indicators of attendance and punctuality to classes, student academic performance and level of commitment and involvement with the process of learning to learn.

4) For user/beneficiary perspective (contribution to the milieu and Interest Groups, such as society, employers and graduates): Contributing to the milieu a graduate with a comprehensive education that contributes professionally to the discipline and growth and development of the society, measured by the level of employability of graduates and the financial compensation that they receive (monthly income from the first and fifth year out of school) as well as evaluated its contribution to the environment in terms of number, quality and evaluation of professional practices carried out, early employability and evaluation to contribution as community services in consulting and business support to entrepreneurs.

For example, the fuzzyBSC could be configured with the following indicators and targets:

1) Organizational Dynamics Dimension or Perspective:

Number of teachers in relation to enrolled students, the goal is to have a higher than average rate of school of the discipline and improve it by 10% annually.

Number of teachers with grades and graduate, maintaining a higher than average rate of school of the discipline with an annual growth by 10%.

2) *Internal Processes Dimension:*

Retention = abandonment of the career by class/total enrolled number by class, the goal is to have below than average rates of school of the discipline and reduce them by 10% per semester.

Repetition = the number of repeaters by class/total number of pupils in the class, maintain a repetition below than average of schools of discipline.

3) *Dimension of customers:*

Level of satisfaction of the learning experience higher than the average of school of the discipline, reducing the gap in annual rates by 10%.

Improve the entry profile reducing gap between the last score accepted with the scoring accepted for the leader school by 10% annually.

Increase enrolled students by 10% over the previous year.

4) *Users/beneficiaries Dimension:*

Employability level higher the average of graduates schools of discipline and reducing the gap in annual rates of 10%.

Wage levels higher the average for graduates of schools of discipline, reducing the gap in annual rates of 10%.

III. SCOPE OF STUDY

The status of certified quality of the learning process of the university system for undergraduate students, according to the Higher Education Council of Chile, are summarized in Table II, which states that only 444 of 7.746 careers (6%), with certified quality. Moreover 4.327 careers (56%) belong to universities with institutional quality certified, there is not binding relationship between the certified quality of the institution and the quality of the learning process specific to each careers, a situation that is being resolved. For post-graduate programs, 1083 programs are offered, of which 79% belongs to such accredited institutions, and 28% of these programs are with certified quality.

TABLE II
RESULTS OF QUALITY CERTIFICATIONS

Units	Total	Certified	Rejected
Universities	65	56	9
Professional Institutes	43	11	1
Technical Learning Centers	74	8	2
Armed Forces Institutions	3	3	3
Careers	7746	444	8
Master Programs	768	179	8
PhD Programs	170	118	5

In this paper, the career under study has a growing labor demand, this means that their income levels have been rising at rates above 10% per year, it is conducted in 94 sites of 33 higher educational establishment, of which 9 have certified quality, it is taught in day and evening modality, with an enrollment of 11,931 students nationwide, it has a low score requirement, their minimum scores of university selection test (PSU), are decreased 39 points in the past five years, although it remains high expectations of achieving employment at egress from 90% with a monthly salary or wages for new graduates of the school not leadership average around \$ 350,000 (500 Euro/month), while those of the leading university exceed \$ 700,000 (1,000 Euro/ month), furthermore, it is considered one of the university careers which has the best return on investment, ie an IRR of 2.8, measured in terms of remuneration of the first 10 years and future income to 65 years, minus the costs of career and in graduating time. The dropout rate in the first year amounts to 30% on average, among the main motivations are the motivational, academic achievement, feelings of inequality by the educational institution of origin and family support.

In the particular case under study, the career has about 20 years of existence with an accumulated historical enrollment of 3,906 students. Of which 1.732 belong to daytime class and 2.174 evening class.

Today the career gives more than 100 courses per semester for its 853 active students.

IV. EXAMPLE OF APPLICATION

As an example, it is only presented the internal process dimension of fuzzyBSC, the data used are from the first semester of the first year, in particular the class entering the career in 2009. The exploratory analysis of these data is performed using the results of the 5 initial subjects that make up the curriculum, these data show the distribution of students per course shown in Table III:

TABLE III
COURSES FOR SUBJECTS OF 1ST SEMESTER.

Subject	Courses	Students
Accounting	4	246
Administration	4	247
Logical Thinking	6	258
Algebra	6	178
English	8	201
Total	28	1.130

When the first semester ended, this class has a good yield 38% on average per subject and 41% of regular yield.

Students generally come from educational establishment of medium-low socio-economic level with a regular-type education, 43% of them correspond to very young students (recent egress), and 39% for adult students from labor world.

From exploratory analysis of data for the results of the first semester of the students entering in 2009, and the opinion of teachers involved in their learning process, who suggesting as an answer to the question "What are the characteristics of students in first year who obtaining good results during the learning process?", they respond as follows:

a) Most students who get better results are very young and came from schools located in a good income township, with a good scoring on the university selection test (PSU).

b) Most students who come from communities of middle income level with low scoring on the PSU, achieve the minimum score needed to pass the course.

c) Usually students entering the career have had better scores on the mathematical PSU than language PSU, which allows them to pass specialty classes, albeit with difficulty.

In general the freshmen, the difficulty to pass first semester courses in decreasing order of difficulty are Mathematics, Accounting, Language and Administration.

From the results obtained by students in the first semester is setting up a database and through an exploratory analysis, the teachers determine the variables that comprise the indicators identified previously by the teachers and the thresholds of the goals, which are the following:

o Whereas 75% of students come from low-educational institutions, we must work to improve the

final average grade and improve it at the rate of 10% per semester.

o Reduce the gap with the entry profile to 5% during the first two weeks of classes.

V. VERIFICATION

To validate the validity of assumptions, the management indicators and targets formulated by teachers for *fuzzyBSC* internal process dimension, we proceed to label the fuzzy variables as passing grade per course completed, PSU score, income level of the home education establishment, and we extracted the no-evident knowledge existing in the databases by extracting fuzzy association rules. Labels were established with the help of experts. This process was developed considering the following:

363 transactions representing the results obtained by students in each of their subjects.

In processing these transactions, it will get the following fuzzy association rules:

507 rules with 2 elements,
2.132 rules with 3 elements,
4.583 rules with 4 elements

Among these 7.222 rules, Table IV shows some very strong rules found.

VI. CONCLUSION

The proposed solution provides to the learning process a measurement and control tool, which involves protagonists with their performance and allows through management indicators determine whether each stage of the process is meeting the standards that the unit has been defined as parameters which ensure its quality and ensure the formation of the egress profile.

This solution reduces or eliminates the uncertainty of educational planning, whose results are usually known at the medium and long term, to change to online control, step by step, stage by stage and it is controlled as the transactions are generated, which allows to know, on the go, the gap between the planned and actual and to make appropriate decisions to correct the process.

Having the possibility of rescuing the useful no-explicit knowledge hidden in the database and extract it using fuzzy data mining techniques, allowing the feedback to maintain force *fuzzyBSC*.

Future work consists of extending the verification process to all dimensions of fuzzyBSC design, through data mining process. Construct and implement the proposed model, considering the necessary interfaces to capture and consolidate data coming from different existing information systems and building up a database which allows to manage online the quality of learning process of undergraduate college, with a repository of historical transactions that periodically allows performing the data mining process that helps maintain in force the management indicators and process goals.

REFERENCES

- [1] R.S. Kaplan and D.P. Norton, "The Balanced Scorecard: Measures that drive performance", Harvard Business Review, 1992.
- [2] F. Bobillo, M. Delgado, J. Gómez-Romero, and E. López, "A semantic fuzzy expert system for a fuzzy balanced scorecard", Expert System with Applications 36, pp. 423-433, 2009.
- [3] R. Fernández, Y. Martínez, N. Velasco, "Indicadores y estándares internacionales de calidad universitaria", Revista Calidad de la Educación 25, pp. 17-29, 2006.
- [4] A. de la Orden, "Evaluación y calidad: análisis de un modelo", Estudios sobre Educación, 16, pp. 17-36, Servicios de Publicaciones de la U. de Navarra, 2009.
- [5] L.A. Zadeh, "Fuzzy Sets", Information and Control. 8, pp. 338-353., 1965.
- [6] B. Kosko, "Fuzzy systems as universal approximators". IEEE Transactions on Computers, 43 (11), pp. 1329-1333, 1994.
- [7] J. Gil Aluja, "Nuevas técnicas para la gestión de empresas. Retos Empresariales", Management Empresa, Universidad de Barcelona, 1997.
- [8] M. Delgado, N. Marín, D. Sánchez, and M.A. Vila, "Fuzzy Association Rules: General Model and Applications", IEEE Transactions on Fuzzy Systems 11(2), pp. 214-225, 2003.
- [9] R. Agrawal, T. Imielinski, and A. Swami. "Mining associations between sets of items in massive databases". In ACM-SIGMOD International Conference on Data, pp. 207-216, 1993.
- [10] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila., "Measuring the Accuracy and Interest of Association rules: A New Framework.", Intelligent Data Analysis 6(3) pp. 221-235, 2002.
- [11] M.J. González and M.E. Cañadas M, "Los indicadores de gestión y el cuadro de mando en las entidades no lucrativas", CIRIEC-ESPAÑA 63 pp. 227-252, 2008.
- [12] L. Arias-Montoya, J.C. Castaño-Benjumea, A.M. Lanzas-Duque, Scientia et Técnica Año XI, 27, 2005, 181-184.

TABLE IV
SOME VERY STRONG RULES

Fuzzy Association Rule	Support %	CF
If SINCEMATFUZZY = 'medium' then SINCELENGFUZZY = 'medium'	75.48	1
If FUZZYINGCOMUNA = 'Low' then SINCELENGFUZZY = 'medium'	72.72	0.028
If AGE = 'VeryYoung' y SINCEMATFUZZY = 'medium' then SINCELENGFUZZY = 'medium'	46	1
If AGE = 'VeryYoung' y FUZZYINGCOMUNA = 'Low' then SINCELENGFUZZY = 'medium'	41	0.1
If FUZZYINGCOMUNA = 'Low' then COMPETENCIAFUZZY = 'NearlyCompetent'	39	0.02
If COMPETENCIAFUZZY = 'NearlyCompetent' y FUZZYINGCOMUNA = 'Low' then SINCELENGFUZZY = 'medium'	36.3	0.3
If AGE = 'VeryYoung' y FUZZYINGCOMUNA = 'Low' y SINCEMATFUZZY = 'medium' then SINCELENGFUZZY = 'medium'	35.8	1
If AGE = 'VeryYoung' y FUZZYINGCOMUNA = 'Low' y SINCELENGFUZZY = 'medium' then SINCEMATFUZZY = 'medium'	35.8	0.48
If FUZZYPSU = 'AlmostIncompetent' then AGE = 'VeryYoung'	29.47	0.22