Real-Time Fuzzy Linguistic Analysis of Anomalies from Medical Monitoring Devices on Data Streams

Javier Medina*
Department of Computer Science. University of Jaen
Jaen, Spain
jmquero@ujaen.es

Macarena Espinilla
Department of Computer Science. University of Jaen
Jaen, Spain
mestevez@ujaen.es

Christopher Nugent
Computer Science Research Institute
Ulster University
UK
cd.nugent@ulster.ac.uk

ABSTRACT

Analysis of data streams generated from devices collecting data from patients, which are monitored within both clinical and home environments, provide useful information for decision making processes. Nevertheless, medical personnel are still required to review and process the data and therefore spend a lot of time and effort to detect situations of concern such as exacerbations with conditions or the occurrence of anomalies in the measurements. In this paper, we propose a methodology for the real-time linguistic analysis of data streams generated from medical monitoring devices based on a rule-based inference engine exploiting a fuzzy linguistic approach. A case study based on health data provided by the Physiological Data Modeling Contest is used to illustrate the proposed methodology and to demonstrate the flexibility to interpret, in a linguistic manner, data streams and the detection of risk situations of interest based on linguistic expressions.

CCS Concepts

•Theory of computation → Streaming models ; •Computing methodologies → Vagueness and fuzzy logic; •Applied computing → Health care information systems;

Keywords

Data streams; Medical monitoring devices; Linguistic modeling; Fuzzy logic; Fuzzy linguistic approach

1. INTRODUCTION

Providing an adequate level of health services requires the analysis of vast amounts of generated from medical monitoring devices [1]. Nevertheless, to measure health related information from patients, in Primary or Secondary care settings or alternatively within a home setting, in an effort to monitor their health status requires, a significant amount of resources. In addition to the resources required, the numbers of persons potentially requiring such a service is also increasing as evidenced in estimations of the percentage of population over 65 rising by up to 15% [2]. Current health-care systems are struggling to manage this population growth and therefore efforts are being made to find new and emerging solutions to manage this challenge.

In an attempt to over this challenge, a methodology to detect anomalies from data streams generated by medical monitoring devices using linguistic terms is proposed. Such a linguistic modeling approach is capable of describing, in a simple and interpretable manner [3], the raw data provided by medical monitoring devices based on the expert knowledge which has been defined by healthcare personnel using linguistic rules.

The proposed methodology is based on the fuzzy linguistic approach and fuzzy logic [4] in conjunction with the use of fuzzy rules that include fuzzy linguistic temporal terms such as *around three days ago* or *around one week ago* as well as fuzzy quantifiers such as *low* or *high*. An inference engine [5] is then used to calculate the matching of these linguistic terms with the data streams to provide a real-time response. Our proposal is based on the fuzzy linguistic approach due to the fact that fuzzy logic is more appropriately suited than traditional logic to deal with concepts which are imperfect, vague and subjective [6] as, for example, *high temperature recently*. In our approach, we aim to process values obtained by medical monitoring devices to linguistic variables, which have a set of linguistic terms in addition to temporal components and the quantifier of these values within this component.

Therefore, the efforts of this contribution have been focused on describing a linguistic methodology to process data streams within a temporal component by using linguistic terms in order to detect anomalies or situation of risks for patients. From the perspective of temporality, fuzzy logic offers an intuitive [7] [8] representation of temporal interval terms by means of fuzzy logic. In addition, we integrate fuzzy quantifiers based on a linguistic ordered scale, which provides a set of ordered quantifiers to evaluate the frequency of a phenomenon with symmetry and balanced properties. As result, this methodology provides richer fuzzy rules, which are adapted to the health expert knowledge, to monitor the data streams in real time. Furthermore, in this

*Corresponding author

.

PervasiveHealth 2016, May 16-19, Cancun, Mexico

DOI 10.4108/eai.16-5-2016.2263877

work, we include a case study where two fuzzy rules detect a situation of risk for a patient from a near-body thermometer monitoring device.

The remainder of the paper is structured as follows: Section 2 briefly reviews the fuzzy linguistic approach. Section 3 proposes a methodology for analyzing data streams based on linguistic the fuzzy linguistic approach, Section 4 presents a case study that follows the proposed methodology and Section 5 presents the conclusions of the contribution madre from this research.

2. FUZZY LINGUISTIC APPROACH

The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables [4]. To do so, it is necessary to choose appropriate linguistic descriptors for the linguistic term set and their semantics.

In the linguistic approach an important parameter to determine is the cardinality of the linguistic term set, which indicates the degree of discrimination given by a term set.

The semantics of the terms are given by membership functions. In this contribution, we shall use as semantics of the linguistic terms trapezoidal membership functions, TS, whose representation is achieved by a 4-tuple (a, b, c, d), where b and c indicate the point in which the membership value is 1, with a and d indicating the left and right limits of the definition domain of the membership function.

The proposed methodology in this paper is based on the fuzzy linguistic approach for analyzing data streams based on expert knowledge of healthcare personnel using the linguistic terms described and a Rule-Based Inference Engine designed.

3. PROPOSED METHODOLOGY

In this Section, we present a methodology for monitoring data streams from health devices using a fuzzy linguistic approach. In Section 3.1, we describe the process of defining fuzzy linguistic terms on single medical monitoring device. In Section 3.2, we detail a proposal of fuzzy quantifiers based on the linguistic scale to quantify the frequency. In Section 3.3, we present a Rule-Based Inference Engine designed for analyzing several data streams.

3.1 Linguistic expressions for data streams

Each data stream s^j can be represented as a set of measures $s^j = \{m_i^j\}$, where each measure is represented by $m_i^j = \{v_i^j, t_i^j\}$; where v_i^j represents a measure that depends on each medical monitoring device, for example, temperature, heart rate, etc.; and t_i^j represents the timestamp, the time at which this measurement was recorded by the health device. In order to model the information, we propose to represent a set of measurements from each medical monitoring device by means of linguistic terms. To achieve this, we propose a self-dependent definition of each component v_i^j and t_i^j and a subsequent method to fuse both linguistic terms.

Firstly, for each data stream s^j, we define a fuzzy linguistic variable, which are all based uniquely on the nature of the medical monitoring device. The general process is described as translating the measures to a new fuzzy set $V = \{V_r; r = 1, ..., n\}$ which represents the linguistic terms. Vr is characterized by a membership function $\mu_{Vr(v_i^j)}$, which is interpreted as the degree of membership of a measure v_i^j in the fuzzy set V_r for each $v_i^j \in m_j^i \in s^j$. For a given linguistic term, we can write $V_r(v_i^j)$ instead of $\mu_{V_r(v_i^j)}$.

Secondly, we associate fuzzy linguistic terms $T = \{T_k; k = 1, ..., m\}$ with the temporal component of the data stream. The degree of this temporal linguistic term is obtained by the temporal membership function and the temporal period that is defined by the distance $\Delta t_i^j = t_0 - t_i^j, t_0 > t_i^j$ from a reference point of time t_0 to the timestamp of the measurement t_i. Each temporal linguistic term T_k relates to the timestamp of the measurement t_i^j to a fuzzy set T_k, which is characterized by a membership function $\mu_{\tilde{T}_k}(t_0 - t_i^j)$. For a given temporal linguistic term, we can write $T_k(\Delta t_i^j)$ instead of $\mu_{\tilde{T}_k}(t_0 - t_i^j)$.

The degree of membership of all measures are aggregated using the t-conorm operator [9] in order to obtain a single degree of fuzzy sets $V_r \cap T_k$ for the given data stream over a period of time:

$$V_r \cup T_k(s^j) = \bigcup_{m_i^j \in s^j} V_r \cap T_k(m_i^j) \in [0, 1] \quad (1)$$

In this contribution, we relate $V_r \cap T_k(s^j)$ to represent the degree of membership between [0,1] of the linguistic term $V_r(v_i^j)$ weighted by the term $T_k(\Delta t_i^j)$ in the data stream. To achieve this, we measure the degree of membership of $V_r \cap T_k(s^j)$ as the fuzzy weighted average [10] related to Eq.(1):

$$V_r \cup T_k(s^j) = \frac{1}{\sum T_k(\Delta t_i^j)} \sum_{m_i^j \in s^j} V_r(v_i^j) \times T_k(\Delta t_i^j) \in [0, 1] \quad (2)$$

In the following example, we present the calculus of *temperature was high for a half day* measured from a thermometer device by means of trapezoidal function (TS), whose representation is achieved by a 4-tuple that represents a function trapezoidal shaped: $TS(a, b, c, d)$. In Figure 1, we present an example of of the linguist term $V_r(v_i^j)$=*temperature is high* weighted by the temporal linguistic term $T_k(t_i^j)$=*since a half day* using Eq. (2) from a data stream over the fuzzy sets $V_1(v_i^j) =$ IsHigh (v_i^j) with $TS\_isHigh = (37.5, 38.5, 41, 41)$ and $T_1(t_i^j) =$ sinceAHalfDay (t_i^j) with $TS\_sinceAHalfDay = (12, 12, 20, 20)$.

$$s^j = \{m_i^j\} = \{v_i^j, \Delta t_i^j\} = \{\{37.5^\circ, 2h\}, \{37.8^\circ, 4h\}, \{38.3^\circ, 6h\}, \{38.6^\circ, 8h\}, \{38.0^\circ, 10h\}, \{38.7^\circ, 12h\}, \{37.9^\circ, 14h\}, \{37.5^\circ, 16h\}, \{38.3^\circ, 18h\}, \{38.6^\circ, 20h\}, \{37.4^\circ, 2h\}, \{37.8^\circ, 24h\}\}$$

$$V_1(v_i^j) = IsHigh(v_i^j) = \{0.33, 0.53, 0.87, 1.0, 0.67, 0.47, 0.6, 0.33, 0.87, 1.0, 0.27, 0.53\}$$

$$T_1(v_i^j) = sinceAHalfDay(t_i^j) = \{1, 1, 1, 1, 1, 0.75, 0.5, 0.25, 0, 0, 0\}, \sum T_k(\Delta t_i^j) = 7.5$$

$$V_r \cup T_k(s^j) = \frac{1}{7.5} \sum \{0.33, 0.53, 0.87, 1.0, 0.67, 0.47, 0.45, 0.17, 0.22, 0, 0, 0\} = 0.63$$

Calculus of degree of the linguistic term *Temperature is high since a half day* (0.63).

3.2 Fuzzy quantifiers to describe the frequency

In this work, we related the linguistic ordered scale to fuzzy quantifiers [11] in order to evaluate the frequency of the linguistic terms V_r and T_k in the data streams. The use of these linguistic terms and scales, allow healthcare personnel to provide richer knowledge in natural language [12, 13] to analyze the data streams. The fuzzy quantifiers are commonly used to aggregate and evaluate a fuzzy set in time series [14].

Firstly, we translate the degree of frequency of Eq. (3) to a membership degree in a linguistic order scale $Q = \{Q_s; s =$

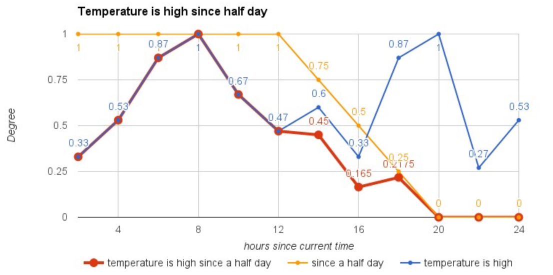

Figure 1: Graphical representation of the linguist term *temperature is high* weighted and the temporal linguistic term *since a half day*

$1, ..., t\}$, where Q_s is a quantifier and the index s represents the order in the scale. To develop this transformation function $Q_s : [0, 1] \rightarrow [0, 1]$, we propose a fuzzy linguistic representation, where each quantifier is described by a symmetric and balanced isosceles triangular membership function in [0,1] and the terms are equispaced, which we refer to as μ_{Q_s}. For example, from a linguistic order scale, where the quantifiers are $Q = \{Q_1, Q_2, Q_3\}$={*lowFrequency, normalFrequency, highFrequency*}, the given membership functions are presented in Figure 2.

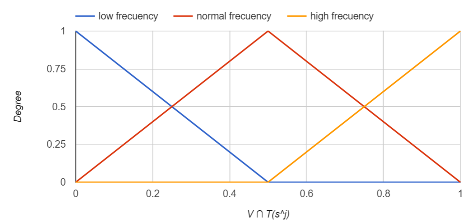

Figure 2: Example of membership functions of a linguistic order scale, where the quantifiers are Q={lowFrequency, normalFrequency, highFrequency}

Secondly, we calculate the aggregated degree of a linguistic quantifier Q_s using the frequency degree of $V_r \cap T_k(s^j)$ and applying an α-cut:

$$\mu_{Q_s}(V_r \cap T_k(s^j), \alpha) = \mu_{Q_s}(V_r \cap T_k(s^j)) > \alpha \qquad (3)$$

In the following example, we present the calculus of the linguistic terms *high temperature since a half day with normal frequency* and *high temperature since a half day with high frequency.*

$$\text{high temperature since a half day}, V_r \cup T_k(s^j) = 0.63$$

$$Q = \{\text{lowFrequency, normalFrequency ,highFrequency}\}, \alpha = 0.5$$

$$\text{high temperature since a half day with normal frequency}$$

$$Q_{normalFrequency} = Q_2, \mu Q_2(0.63, 0.5) = Q_2(0.63) > 0.5 = 0.74 > 0.5 = \mathbf{0.74}$$

$$\text{high temperature since a half day with high frequency}$$

$$Q_{highFrequency} = Q_3, \mu Q_3(0.63, 0.5) = Q_3(0.63) > 0.5 = 0.24 > 0.5 = \mathbf{0.0}$$

3.3 Rule-Based Inference Engines

In order to carry out the process of analyzing data streams of patients, we propose a Knowledge-Driven Approach (KDA), which makes use of rich domain-specific expert knowledge, based on a Rule-Based Inference Engine. The Inference Engine processes linguistic terms from one or multiple medical monitoring devices by means of temporal fuzzy rules and linguistic ordered quantifiers as previously described. The proposed Rule-Based Inference is based on a collection of fuzzy logic rules in the form of IF-THEN statements, where the proposed linguistic terms and quantifiers are included. The antecedent A_i are represented as:

A_i: s^j THERE IS V_k WHEN T_k WITH Q_s

The degree of A_i, $\mu_{Q_s}(A_i)$, is computed as described in Eq.(4). The consequents, represented as C_j, are context-dependent linguistic terms that define a state of the patient or alert, for example, *patient is with fever* or *notify nurse supervisor*. The full structure of the rule is:

$R_{j_{\alpha_j}}$: IF A_1 and A_2 and A_n THEN C_j

The inference engine will determine the degree of matching of antecedents $\{A_1, ..., A_n\}$ from multiple medical monitoring devices aggregating them with the logical $\cap$ operator:

$$\mu_Q(C_j) = \mu_Q(R_j) = \bigcap(\mu_Q(A_i)), \forall A_i \in R_j \qquad (4)$$

where $\mu_{Q_s}(R_j)$ is the degree of matching of all antecedents of the rule R_j which is assigned to consequent C_j. Finally, the output of the Rule-Based Inference Engine returns the degree of matching of all consequents $\mu_C = \{\mu_{C_1}, ..., \mu_{C_n}\}$, together with those antecedents which have matched the rules and which compose the explanation of the rule activations.

4. CASE STUDY

A case study based on health data provided by the Physiological Data Modeling Contest (PDMC) [15] is presented in this Section. The data provided by the PDMC consists of physiological sessions, where each minute of the session is based on nine types of medical monitoring devices. The data was collected from subjects using BodyMedia medical monitoring devices. We have considered the data based on user session id 2733, and sensor row 5. This session depicts over 18 hours of measurements, recorded as a series of one minute recordings (1132 instances), where the temperature of the patient rises at the end of the session (refer to Figure 3 (A)).

First, related to this session, we have defined an example of linguistic analysis based on the methodology proposed in the previous Sections, where the medical monitoring device is a near-body thermometer, a linguistic term is related to *high temperature*, two linguistic terms evaluate two different temporal components: *recently* and *about 6 hours*. In addition, we define two IF-THEN rules and a linguistic order scale where we apply health expert knowledge to analyze the data stream in real time:

- s^1=temperature
- $V = V_1$={isHigh} with $TS_{isHigh} = (0, 32, 34, 34)$.
- $T = T_1, T_2$={recently,about6Hours} with $TS_{recently}$=(0,2,4,4) and $TS_{about6Hours}$=(0,6,8,8).
- $Q = \{lowFrequency, normalFrequency, highFrequency\}$ with $TS_{lowFrequency} = (0, 0, 0, 0.5)$, $TS_{normalFrequency} = (0, 0.5, 0.5, 1)$ and $T_{highFrequency} = (0.5, 1, 1, 1)$.
- $\alpha = 0.5$ as a compressive fuzzy threshold to eliminate noise and uncertainty.
- R_1*: IF temperature IS of high WHEN recently WITH high frequency THEN send supervisor a sms.*
- R_2*: IF temperature IS of high WHEN since about 6 hours WITH normal frequency THEN provide antipyretic.*

Using these linguistic terms and the data from PDMC, we have calculated $V_1 \cap T_1$ and $V_1 \cap T_2$ (refer to Figure 3(B)) over the session time with the proposed Eq. (2).

The rule R_1 analyzes the urgent case when there is high temperature in a short period of time. The rule R_2 represents a normal situation of high temperature due to an antipyretic. In Figure 3(C) we show the real-time rule evaluation by the Inference Engine proposed in Section 3.3 together with the suggested actions for healthcare personnel based on the rules being triggered.

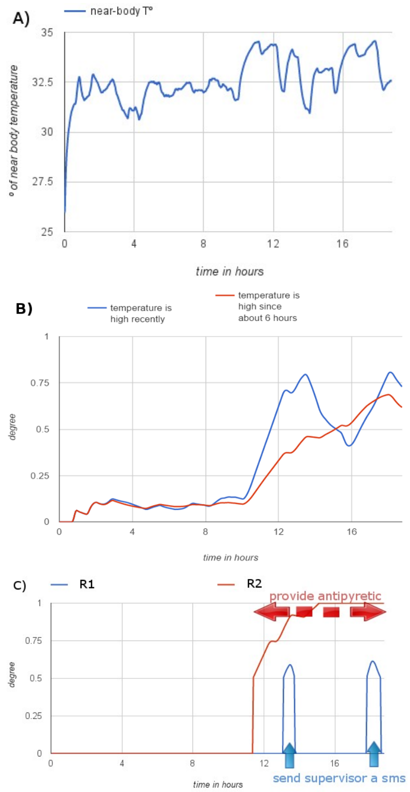

Figure 3: A) Evolution in session time of near body temperature from PDMC (session 2733) B) Calculus of degree of the linguistic terms *normal frequency of high temperature since a half day* and *high frequency of high temperature since a half day*. C) Evolution in session time of rule evaluation and suggested actions related to consequents.

5. CONCLUSIONS

In this contribution, we have proposed the use of a fuzzy linguistic approach as a methodology for analyzing anomalies and situation of risk from sensor streams processed in real time based on a rule-based inference engine. The main advantage of this proposal is that the proposed fuzzy temporal terms and fuzzy quantifiers. This make them very flexible to model the information provided by healthcare personnel in a linguistic manner by means fuzzy rules. A case study based on real data from the PDMC has been developed to convey the proposed concepts. In future works, we will report the results of integrating multi sensor streams with a fuzzy linguistic approach using wearable devices within a health context.

Acknowledgments

This contribution has been supported by research project UJA2014/06/14 and Invest Northern Ireland is acknowledged for supporting this project under the Competence Centre Programme Grant RD0513853 - Connected Health Innovation Centre.

6. REFERENCES

[1] Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. Jama, 309(13), 1351-1352.

[2] United Nations, Department of Economic and Social Affairs, Population Division (2013). World Population Ageing 2013. ST/ESA/SER.A/348.
Schuldt, H., & Brettlecker, G. (2003). Sensor data stream processing in health monitoring.

[3] Mikut, R., Jakel, J., & Groll, L. (2005). Interpretability issues in data-based learning of fuzzy systems. Fuzzy Sets and Systems, 150(2), 179-197.

[4] Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.

[5] Togai, M., & Watanabe, H. (1986). An inference engine for real-time approximate reasoning: Toward an expert on a chip. IEEE Expert, 1(3), 55-62.

[6] Stover, J. A., Hall, D. L., & Gibson, R. E. (1996). A fuzzy-logic architecture for autonomous multisensor data fusion. Industrial Electronics, IEEE Transactions on, 43(3), 403-410.

[7] Ohlbach, H. J. (2004, July). Relations between fuzzy time intervals. InTemporal Representation and Reasoning, 2004. TIME 2004. Proceedings. 11th International Symposium on (pp. 44-51). IEEE

[8] Ros, M., Cuellar, M. P., Delgado, M., & Vila, A. (2013). Online recognition of human activities and adaptation to habit changes by means of learning automata and fuzzy temporal windows. Information Sciences, 220, 86-101.

[9] Gupta, M. M., & Qi, J. (1991). Theory of T-norms and fuzzy inference methods. Fuzzy sets and systems, 40(3), 431-450.

[10] Dong, W. M., & Wong, F. S. (1987). Fuzzy weighted averages and implementation of the extension principle. Fuzzy sets and systems, 21(2), 183-199.

[11] Glockner, I., & Knoll, A. (2001). Fuzzy quantifiers: A natural language technique for data fusion. In Proceedings of the Fourth International Conference on Information Fusion (Fusion 2001).

[12] Hu, H. Y., Lee, Y. C., & Yen, T. M. (2010). Service quality gaps analysis based on Fuzzy linguistic SERVQUAL with a case study in hospital out-patient services. The TQM Journal, 22(5), 499-515.

[13] Li, Q. (2013). A novel Likert scale based on fuzzy sets theory. Expert Systems with Applications, 40(5), 1609-1618.

[14] Sanchez, D., Bugarin, A., Nicolas, M., & Trivino, G. (2015). Fuzzy Knowledge Representation for Linguistic Description of Time Series. 16th World Congress of the IFSA- 9th Conference of the EUSFLAT, Gijon (Spain), Atlantis Press, pp. 1346-1353.

[15] Andre, D., & Stone, P. (2004). Physiological data modeling contest. In A ICML (Vol. 4).