Computing protoforms from high-rate sensor streams

4 authors:

J. Medina
University of Granada
119 PUBLICATIONS 1,411 CITATIONS

Macarena Espinilla
University of Jaén
144 PUBLICATIONS 2,314 CITATIONS

Luis Martinez
University of Jaén
614 PUBLICATIONS 29,605 CITATIONS

Sixto Enrique Campaña Bastidas
Universidad Nacional Abierta y a Distancia
35 PUBLICATIONS 85 CITATIONS

Computing protoforms from high-rate sensor streams

J. Medina[1], M. Espinilla[1], S. Campana[2], L. Martinez[1]

[1] *Department of Computer Science. University of Jaen, Spain*
jmquero@ujaen.es, mestevez@ujaen.es, luis.martinez@ujaen.es
[2] *Universidad Nacional Abierta y a Distancia - UNAD, Colombia*
sixto.campana@unad.edu.co

In the context of the Internet of Things (IoT) many sensors have been enhanced with respect to their ordinary role to be proactive and collaborative with other devices. This fact provokes a description of the world and the human activity as we have never collected [1].

The data streams provided by the set of sensors is continuously increasing. Generally, these data streams are just analyzed and partially stored. In order to interpret the information of sensor streams is critical developing linguistic descriptions on natural language in a comprehensiveness and interpretability way, depending of the application domain.

In this contribution, we introduces a methodology to compute linguistic descriptions of high-rate sensor streams that considers relative quantifiers on a temporal component by means of *protoforms* [2], which are an useful knowledge model for summarization [3], quantification [4] or in time series [5]. In this work, we propose protoforms in the shape of s^j *is* $Q_s^j\ V_r^j\ T_k$, to represent the linguistic terms V_k^j which describe the sensor stream s^j with the relative quantifier Q_s^j in a temporal terms T_k.

An example of linguistic description computed by the proposed methodology from a data stream of a pulsometer is as follows: *the most of heart rate measurements are high recently*.

Methodology Each sensor data stream s^j is represented as a set of measures $s^j = \{m_i^j\}$, where each measure is composed by $m_i^j = \{v_i^j, t_i^j\}$; where v_i^j represents a value and t_i^j represents the related timestamp of v_i^j provided by the sensor. For each sensor data stream s^j, we describe a fuzzy linguistic term set $V^j = \{V_r^j; r = 1,...,n\}$ which represents the linguistic terms of the sensor values. Secondly, we associate a set of fuzzy linguistic temporal terms $T^j = \{T_k^j; k = 1,...,m\}$ with the temporal component of the data stream. For sake of simplicity, we write $V_r^j(v_i^j)$ and $T_k^j(t_i^j)$ instead of $\mu_{\tilde{V_r^j}}(v_i^j)$ and $\mu_{\tilde{T_k^j}}(t_i^j)$.

Our methodology proposes to compute the degree of relevance of a linguistic value term $V_k^j(v_i^j)$ in a temporal term $T_k^j(\Delta t_i^j)$ in the sensor streams, being $\Delta t_i^j = |t_0 -_i^j|$ and t_0 the current time. To do so, we propose the use of aggregation operations [6] based on uninorms using t-norm t-conorm [7], respectivelly:

$$V_r^j \cap T_k^j(s^j) = \bigcup_{m_i^j \in s^j} V_r^j(v_i^j) \bigcap T_k^j(t_i^j) \in [0,1] \tag{1}$$

This contribution has been supported by Research Projects UJA2014/06/14

The Eq. (1) provides the degree of the protoform instances in the shape of s^j *is* V_r^j T_k^j, for example, *heart rate measurements are low recently*. In order to introduce relative quantifiers in the protoform s^j *is* Q_s^j V_r^j T_k^j, each quantifier Q_s^j is described by a transformation function $\mu_{Q_s^j} : [0,1] \rightarrow [0,1]$, translating the degree from Eq. (1) to a quantified aggregated degree $Q_s^j(s^j) = (V_r^j \cap T_k^j(s^j))$.

We highlight that protoforms require an accurate assessment to express properly their semantic and granularity. For example, to aggregate measures of high sample rates where data are frequent and spaced, we propose the fuzzy weighted average [8], which represents the frequency degree of the linguist term V_r^j weighted by the term T_k^j in the sensor stream.

$$Q_s^j(s^j) = V_r^j \cap T_k^j(s^j) = \frac{\sum_{m_i^j \in s^j} (V_r^j(v_i^j) \cdot T_k^j(\Delta t_i^j))}{\sum_{m_i^j \in s^j} T_k^j(\Delta t_i^j)} \in [0,1] \tag{2}$$

References

[1] Bogue, R. (2014). Towards the trillion sensors market. Sensor Review, 34(2), 137-142.

[2] Zadeh, L. A. (2006). Generalized theory of uncertainty (GTU)principal concepts and ideas. Computational Statistics & Data Analysis, 51(1), 15-46.

[3] Kacprzyk, J., & Zadrony, S. (2005). Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools. Information Sciences, 173(4), 281-304.

[4] Zadeh, L. A. (1983). A computational approach to fuzzy quantifiers in natural languages. Computers & Mathematics with applications, 9(1), 149-184.

[5] Sanchez, A. B. N. M. D., & Trivino, G. (2015). Fuzzy Knowledge Representation for Linguistic Description of Time Series.16th World Congress of the International Fuzzy Systems Association - 9th Conference of the European Society for Fuzzy Logic and Technology, Gijn (Spain), Atlantis Press, pp. 1346-1353.

[6] Medina, J., Espinilla, M., Moya, F. & Nugent, C. (2015). Activity recognition by means of rule-based inference engine based on fuzzy linguistic approach. The 13th Scandinaviane Conference on Artificial Intelligence. Sweden.

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

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