2009 Ninth International Conference on Intelligent Systems Design and Applications

Automatically generated linguistic summaries of energy consumption data

Albert van der Heide
Cognitive Computing: Computing with Perceptions
European Centre for Soft Computing
Mieres, Spain
albert.vdheide@softcomputing.es

Gracián Triviño
Cognitive Computing: Computing with Perceptions
European Centre for Soft Computing
Mieres, Spain
gracian.trivino@softcomputing.es

***Abstract*—In this paper a method is described to automatically generate linguistic summaries of real world time series data provided by a utility company. The methodology involves the following main steps: partitioning of time series into fuzzy intervals, calculation of statistical indicators for the partitions, generation of summarising sentences and determination of the truth-fullness of these sentences, and finally selection of relevant sentences from the generated set of sentences.**

***Keywords*-summerization; fuzzy logic; quality in agreement; relevance;**

I. Introduction

The collection and storage of data has become relatively easy. This wealth of information will need to be processed and interpreted to be useful, which has raised interests in fields such as data mining, data warehousing and knowledge discovery, as well as data summarization in general [1], [2], [3].

Typically knowledge discovery in databases is designed to extract very precise and hidden information from the data rather than providing a global view of the whole database [4], using a representation unintelligible for the user with a sharp defined conceptual framework.

In everyday language and interaction no sharp concepts are used and the boundaries of concepts become vague, especially when considering a large group of non-expert individuals. Fuzzy logic [5], [6] provides a tool that can model this vagueness of everyday language concepts. Furthermore fuzzy logic makes it possible to analyse and describe complex systems in linguistic terms instead of numerical values [7].

Significant work has been done in the field of linguistically summarizing databases with use of fuzzy logic, for example by Kacprzyk and Zadrożny [8], [9] with a particular focus on so called protoforms, and by Yager [10].

Another field in which the use of fuzzy logic is being explored is for generating linguistic summaries of time-series data [11].

Linguistic summaries are meant to be a general, human-consistent description of data sets, which capture the core trends of the data. These summaries are not meant to be a replacement for classical statistical analysis but rather an alternative means of representing the data focussed on quick human understandability and interpretability; the summaries are brief descriptions of trends in the data stated in natural language.

An implicit advantage of interest to the company is the potential of linguistic summaries to give the impression of personal attention to the clients over numerically presented data; as if an expert analysed the consumption data and generated a report.

The work presented here is an example of application of softcomputing techniques for summarization of energy consumption data of households, to be send as part of the invoices to be send to clients.

A relational database of basic sentences is constructed on the bases of energy consumption data. From this database summaries are constructed in the form of *'Q objects in R are S'* [12], where Q is the quantity in agreement, R a data collection and S the summarizer.

As there are too many summarising sentences that can be produced for a particular household a selection has to be made. The selection is performed based on (1) the validity (truthfulness) of sentences and (2) on a relevance measure. The relevance measure is based on sentence validity values from all households.

The structure of this contribution is as follows. In section II an overview is given of the problem and the solution methodology on which the subsequent sections III, IV, V and VI elaborate. Section VII describes the system output and finally conclusions are made in section VIII.

II. The problem and the solution methodology

In this paper a methodology is described that has been used to summarise time series data collected by the utility company HC Energía, located in the province Asturias, Northern Spain [13]. HC Energía occupies 94% of the total energy market in Asturias. The data collected contains a description of the electrical power consumption of the company's clients, e.g. households. In a 35 of these households watt-meters have been installed that register the quarterhourly cumulative energy consumption.

HC Energía has introduced a variable pricing scheme, which means households pay less for their energy consumption during low consumption hours, e.g. during the early morning

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.8

and night. The company would like to use the collected data to give a summary to the clients about their consumption and, if reasonable, give advice to them on how to change consumption behavior to reduce money spending.

Hence, our project has two main elements which are the automatic summarization of the data and secondly the automatic generation of advice to the clients based on the data. This contribution collects the work done on the first part, while another group of investigators[1] used graph mining techniques to research the possibility to generate recommendations for the households.

The summaries are to be send to households every two months (in Spain, a bimonthly invoice system is established for billing) and need to satisfy some criteria. First of all the summary will need to be understandable by all clients, e.g. undependable upon social group, education, etc. This requisite demands the use of every day natural language and terminology. Secondly the summary will need to give the impression to the clients as if it had been generated by a human expert, e.g. only present 'relevant' data to the clients.

In order to meet the criteria a methodology has been followed involving a number of key elements.

a) selection of meaningful parts of the data and basic linguistic labelling (section III, III-A and III-B)
b) use of statistics to (numerically) summarise parts (section IV and IV-A)
c) design of comparative linguistic labels and their membership functions (section IV-B, IV-B1 and IV-B2)
d) summarization over sets of basic descriptions; a database summary (section V)
e) determining the relevance of all possible database summaries (section VI)

The data used in research has been collected from 35 households in a period ranging roughly from June 2008 to September 2008. (The exact times of the first and last available measurements differ for the clients; the earliest measurement dates from June 21, the latest from September 19.) For summary generation always a subset of all available data was used, namely 70 days of data (70 days is approximately 2 months, which coincides with frequency invoices are send to the households.)

For reasons of privacy the scales in plots have occasionally been omitted when deemed necessary.

III. Descriptional atoms

One of the criteria the summary that will finally be presented to the households must fulfil is that it needs to explain data using everyday language terms. For example, when the summary makes reference to a period in time, terms such as 'on Mondays' or 'in the morning' are preferred above terms such as '08:30 hours until 10:30 hours'.

[1]Research led by Dr. Christian Borgelt, Principal Researcher of the Research Unit "Intelligent Data Analysis and Graphical Models" at the European Centre for Soft Computing.

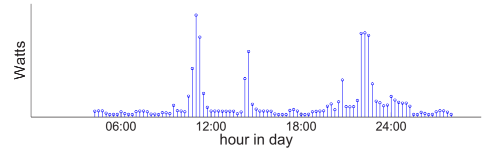

Figure 1. A 'day' of consumption from a household ($4 * 24 = 96$ data points). Typically households have low varying base consumption and short bursts of somewhat erratic high consumption.

However, these natural language terms need to correspond with the interpretation given by the clients. In the next subsections the motivation for the chosen descriptional atoms referring to 'time' is given.

A. A day of data

In order to make summaries using **terms to denote time/day information**, e.g. 'on Mondays' or 'on working days', one needs to define the hours that are part of a day. Officially days run from 00:00 hours until 24:00 hours, however this does not correspond to human use of the term 'day'. For example, when a person is communicating the sentence 'yesterday I saw a program on tv' it is quite possible that the program was seen at 01:00 hours (which officially would refer to today). What this illustrates is that in human interaction 'a day' is roughly defined by when a person wakes up until when that person goes to sleep.

While analysing the data of the households it became clear that there are general reappearing patterns in consumption, clearly showing a 24 hour consumption cycle. This cycle typically shows one or several short periods of highly fluctuating consumption and one long period of continuous low level consumption. Usually the long period of low level consumption coincides with the late-night hours, roughly from 2 o'clock until 6 o'clock. A deduction was made that this period is indicative for when the people in a household are sleeping.

It was decided that days should be separated from one another by the periods in which clients sleep, and therefore a day is defined as starting at 04:00 hours until 04:00 hours the next day (all households in the data have low flat consumption at 04:00 hours. Note however that it is quite unlikely that this will be true for all Spanish households.) See figure 1 for an example of 'a day' of energy consumption data of a particular household. Notice that, four different acquisitions (one each 15 minutes) are taken for each hour representing the cumulative energy consumption within the 15 minute interval.

B. Parts of a day

Analysis of the data can reveal some general properties of the data. Several households show a clear peak in the morning. However, this peak is located at different times, likely due to the different times at which people wake up. Often there is also an evening peak, even though this is much wider and less clearly defined. Some customers also show a lunchtime peak, again less pronounced and less clearly defined as the morning peak.

The main problem with these peaks is that their location varies from client to client and, to some degree, also for a particular client depending on the day of the week.

The energy consumption data displays a form of seasonality as the pattern more or less repeats itself each day although this pattern may drift or change in amplitude over time. Note that these types of patterns are present in many domains as for example in the study of economic data and business cycles, or weather patterns.

A natural way to summarise days of data is using crisp terms, e.g. '*from 20:00 hours until 22:00 hours* the energy consumption is above average', or in less precise terms, e.g. '*during the evening* the energy consumption was above average'.

Within the context of this project it was decided that the second type of description would provide the basis for a more natural and flexible summary. Five basic parts of day are identified, namely, 'morning', 'midday', 'afternoon', 'evening' and 'night', as these 'labels' are already present in Natural Language.

A problem with the use of this terminology is that the terms do not have a universal interpretation; what is morning for one person is already considered midday for another. However, there does exist a general consensus in a population as a whole as to what hours are obviously part of 'the morning' and what hours are not (furthermore, an individual could have a particular interpretation of 'morning', because he/she starts the day at 12:00 hours, but still this person would need to conform to the general consensus in communication about the meaning of 'morning' in order to be understood). It is this consensus-meaning that is modelled.

As the concepts 'morning', 'midday', 'afternoon', 'evening', and 'night' are not clearly bounded in time, fuzzy logic can be an excellent tool to model this [14], [15], [16]. Fuzzy logic makes it possible to define intervals that overlap, where consumption values contribute to both intervals. Therefore, when morning consumption peaks partly fall outside the 'crisp' morning area, these peaks still fall to some extent within the 'morning'.

The definition of the parts of a day is as depicted in figure 2. The partitioning has been done by a domain expert, i.a. a native Spaniard living in Spain. Note that the partitioning takes into account a specific cultural context; the intended public for this particular partitioning is Spanish. However, if the audience would be, for example, a Dutch audience the partitioning would most likely be shifted a couple of hours to the left.

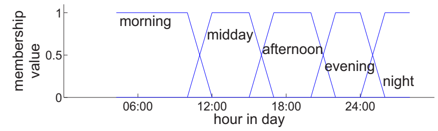

Figure 2. A day of data is divided into five parts using five linguistic terms with their corresponding fuzzy membership functions. The hours within a part coincide roughly with the concepts they are labeled with; 'morning', 'midday', 'afternoon', 'evening', and 'night'.

IV. Summarizing data

There are several ways to summarize data and which option to choose depends highly upon the information one wishes to convey and on what information is considered as relevant. In our work it is decided that information about energy consumption is to be conveyed in terminology such as 'high consumption', 'medium consumption', and 'low consumption'. At this level of detail the trade-off between preciseness of information and its significance (principle of incompatibility [17]) is considered maximised for the majority of clients for this particular application.

In order to be able to decide if a particular consumption is 'high' it is necessary to have a reference to compare to. Comparisons have to be made relative to the household itself as different households contain different numbers of people, of different ages and with different habits. To compare consumption in a part-of-day with other part-of-day for a particular household our work uses a statistical value; the average consumption within a fuzzy interval (see section IV-A). Using this statistical indicator two sets of fuzzy functions are defined, as will be explained in section IV-B.

A. Statistical indicator

To summarize (numerically) the data in a particular part-of-day, e.g. the data characterized by the 'morning' linguistic term, statistical indicators can be used. In this work the average consumption value of a partition is used;

$$\bar{c} = \frac{\sum_{i=1}^{n} c_i u_i}{\sum_{i=1}^{n} u_i} \tag{1}$$

A day of data consists of n samples. c_i denotes the cumulative household consumption for the i^{th} 15 minutes interval. u_i denotes the corresponding membership value for the particular part-of-day. The statistical indicator $\bar{c}$ is used as the basis for comparisons.

B. Comparison and Fuzzification

There are many ways of summarising data in sentences and all of the possible sentences have a particular semantics attached to them. In this work two types of comparisons between day-parts are used as the basis for the semantic sentences:

1) Comparison of a part-of-day with another part-of-day. E.g. "During the night the energy consumption is *higher than* in the morning."
2) Comparison of a part-of-day with a set of day-parts. E.g. "During the night the energy consumption is *low* (with respect to other parts of the day)."

The terms 'higher than' and 'low' are called **summarizers** [12] of the data.

It should be noted that sentences have semantics within a **frame of reference**. In order for a statement 'A is *higher than* B' to be meaningful not only does A have to be of higher value then B, but additionally this difference needs to be significant within a larger domain.

1) Part-of-day compared to part-of-day: In order to be able to meaningfully compare two average consumption values for part-of-days a frame of reference has to be established. To model this frame of reference the **variance in average consumption within a period** is used. When statements about a part of the day are to be generated it is required to at least take into account the variance within the day, which in our work has been used as the frame of reference. Likewise statements using comparisons between days of data require taking into account the variance in values within a week (or more).

The variance in average consumption within a period is calculated straightforwardly by means of the following expression:

$$\bar{c}_{var} = \max_{p=1}^{m}(\bar{c}_p) - \min_{p=1}^{m}(\bar{c}_p) \tag{2}$$

where $\bar{c}_p$ is the p^{th} average consumption in the set of m parts in the used frame of reference. E.g. in this work the frame of reference is 'a day', which has five parts (m), and for each part an average consumption $\bar{c}_p$ is known.

Using the value $\bar{c}_{var}$ fuzzy partitions can be defined representing when A is 'lower than', 'equal to' and 'higher than' B. The membership functions can be defined with use of a variable b:

$$b = \frac{1}{3}\bar{c}_{var} \tag{3}$$

The value of b has been determined by abstracting over the visual inspections of data and reports made by domain experts. Still, the value of b must be interpreted as a heuristic, as there is much variability among clients in consumption patterns.

Using the functions as depicted in figure 3 the comparison between two day-parts can be translated directly into a basic sentence. With five part-of-days and three summarizers there are $(4+3+2+1)*3 = 30$ basic sentences which can be generated (four comparisons of first day-part with following day-parts, plus three new comparisons of second day-part with following day-parts, etc. times three possible labels).

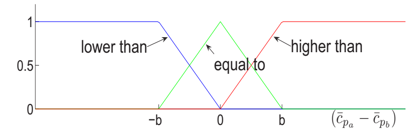

Figure 3. Fuzzy partitions used in the comparison between two average consumption values, $\bar{c}_{p_a}$ and $\bar{c}_{p_b}$.

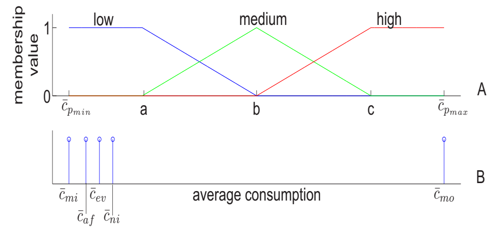

Figure 4. A: The membership functions for labels 'low', 'medium' and 'high' based on average consumption in parts of the day.
B: An example of labelling of for a particular day. The average consumption of 'midday' ($\bar{c}_{mi}$), 'afternoon' ($\bar{c}_{af}$), 'evening' ($\bar{c}_{ev}$), and 'night' ($\bar{c}_{ni}$) fall within the bounds of membership function low, and are consequently labeled as 'low consumption'. The 'morning' ($\bar{c}_{mo}$) is labeled as 'high'.

2) Part-of-day compared to a set of part-of-days: In order to be able to make statements such as 'in the morning the consumption is *high*' also a frame of reference to compare to has to be taken into account. To make explicit the semantic frame of reference implied by the sentence this sentence is restated into a semantically similar one, namely, 'in the morning the consumption is high *when compared with the rest of the day*'.

The frame of reference has been established using $\bar{c}_{var}$ based upon all $\bar{c}_p$ values within a day, as in equation 2. The membership function used to represent the concepts 'high', 'medium' and 'low' are defined using three points (see figure 4):

$$\begin{aligned} a &= \bar{c}_{min} + 0.2\bar{c}_{var} \\ b &= \bar{c}_{min} + 0.5\bar{c}_{var} \\ c &= \bar{c}_{min} + 0.8\bar{c}_{var} \end{aligned} \tag{4}$$

The choice for these specific values a, b, and c has been made pragmatically and are by no means considered definite. An example of the results obtained by this particular partitioning can be seen in figure 4 and translate directly into 15 basic sentences (5 day-parts times 3 possible labels).

V. Quality in Agreement

At this point a number of basic summaries is available for each client, 30 sentences comparing part-of-day to part-of-day and 15 sentences comparing a part-of-day to a set of day-parts. All of these summaries have a membership value or **truthfulness value** assigned to them.

However, the summary to be generated will need to take into account a database of two months data and not just a day. Yager [12] developed a measure to determine the **quality in agreement** of basic sentences with a database of sentences.

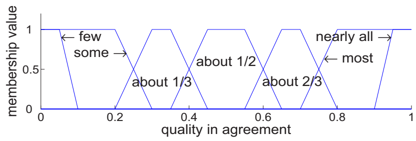

Figure 5. Labels based on the quality-in-agreement value. The membership value can be interpreted as the validity value related to a linguistic summary.

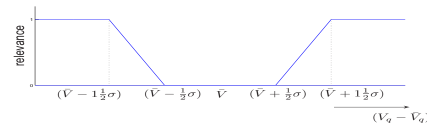

Figure 6. Relevance of sentences is defined by a membership function defined over the distance between the validity of a sentence for a household k, V_q, and the average validity of this sentence for all households, $\bar{V}_q$.

Every basic sentence has a truthfulness attached to it, s_{ij} where i is the sentence number and j is the day, week or month, corresponding to the frame of reference. In our work there are $(30+15)$ possible basic sentences (i) per day (j) in a set of approximately 70 days (n). The quality in agreement of sentence i with the database is given by:

$$Q_i = \frac{1}{n} \sum_{j=1}^{n} s_{ij} \tag{5}$$

This value can be understood intuitively, e.g. if a basic sentence 'The consumption is high in the afternoon' always has a truthfulness of 1 then the quality in agreement is 1. In other words, '*for all days* the consumption is high in the afternoon'.

This quality in agreement can be mapped easily to meaningful linguistic labels; the value Q is fuzzified and labels are attached as depicted in figure 5. The membership value to the various linguistic quantifiers is interpreted as the **validity of a linguistic summary**. Note that there are overlapping fuzzy intervals, e.g. 'few' with 'some' and 'most' with 'nearly all'. When for both intervals a high membership value is obtained then the more restrictive interval provides the best label.

Note that the addition of quantifiers makes it possible to summarise over a set of n days of data, effectively summarising n-days of i basic sentences into $7 * i$ quantified basic sentences, which results in $(4+3+2+1)*3*7 = 210$ possible sentences in case one (section IV-B1) and $5 * 3 * 7 = 105$ possible sentences in case two (section IV-B2). All of these sentences have a validity value V_q attached to them.

VI. Relevance

Many of the quantified basic sentences that can be generated have a high validity value. However, not all of these sentences are of interest to the client. For example the sentence 'almost all days your consumption in the early morning is low' is not interesting as the client is always sleeping at this time and expecting a low energy consumption.

In general, when creating summaries of data from databases the selection of what information is relevant and what not is often problematic. Learning algorithms tend to extract association rules with very large support which are not informative for the user.

This project uses a heuristic to determine relevance: if a particular sentence has a high validity for all households, then that sentence most likely does not contain any remarkable information. If a particular sentence has a high validity value for this household but not for any other, then the sentence contains relevant information.

For every sentence q the average validity $\bar{V}_q$ and standard deviation $\sigma_{\bar{V}_q}$ over all households is calculated. With these values a relevance function R_q is defined for client k using the distance $|V_{q_k} - \bar{V}_q|$, as shown in figure 6 (note that the heuristic will not function well when households have an unusual energy consumption pattern, for example people living and working during the night. However, one could still argue that an expert would also remark this information when looking at the data, and that therefore this information should have a high relevance value.)

VII. Output sentences

There are two main types of sentences that can be generated, 210 possible sentences in for case one (section IV-B1) and 105 possible sentences in case two (section IV-B2). Not all of these sentences should be included in the final summary. To select the sentences to present in the report threshold values have been used based on two variables, namely;

- Validity; only sentences with at least 0.5 validity are selected. This measure expresses the support a sentence has in the database of sentences.
- Relevance; only sentences with a relevance value of at least 0.5 are selected. This measure expresses the distinctiveness of a sentence over all households.

The establishment of the specific threshold values is motivated by the number of output sentences the program is to generate, which has direct influence on the length of the invoice summaries. With the currently established values an average of $3,3$ sentences of type one is generated and $1,9$ sentences of type two. Note that for 4 of the 35 households no sentences of type one were generated. The same was the case for 6 of the 35 households of type two. The reason for this is that these households matched the average household consumption too much.

Examples of the two types of output sentences are given in table I. Templates have been used to generate the sentences.

VIII. Conclusion

The methodology used is straightforward and easily implemented, utilizing standard statistics and fuzzy logic. What is not clear however is how to parameterize all

Table I
EXAMPLE OF OUTPUT SENTENCES FOR A PARTICULAR HOUSEHOLD, 5 SENTENCES OF TYPE ONE AND 2 OF TYPE TWO. LINGUISTIC LABELS ARE PRINTED IN SMALL-CAPS.

validity	relevance	sentence generated
0.831	1.000	ABOUT TWO THIRDS of the days the consumption in the MORNINGs is LOWER THAN the consumption in the AFTERNOONs.
0.920	0.567	MOST of the days the consumption in the MORNINGs is lower than the consumption in the EVENINGs.
0.999	1.000	ABOUT TWO THIRDS of the days the consumption in the MIDDAYs is LOWER THAN the consumption in the EVENINGs.
1.000	1.000	ABOUT TWO THIRDS of the days the consumption in the AFTERNOONs is LOWER THAN the consumption in the EVENINGs.
0.517	0.670	MOST of the days the consumption in the AFTERNOONs is HIGHER THAN the consumption in the NIGHTs.
1.000	1.000	ABOUT TWO THIRDS of the days your consumption is LOW during the MORNINGs.
0.819	1.000	MOST of the days your consumption is HIGH during the EVENINGs.

fuzzy functions, although the success of an implementation depends heavily on this. The functions need to fit seamlessly with the meaning the general public has for the related linguistic labels. In the current work there has been a high dependence upon expert knowledge for the tuning of the parameters. Other solutions could be to automatically learn / tune these parameters by means of, for example, genetic algorithms [18]. Although the current methodology is judged as practical and efficient, it is expected that as the complexity of the membership functions and their number increases parameterization will become more of an issue.

The produced sentences should be easily interpretable by the clients. However, evaluation if this was the case could not be made directly, due to various practical reasons. Therefore the evaluation of the results has been done by experts from the utility company and investigating researchers. Data-plots of household consumption are compared with the generated NL descriptions. Using this heuristic the results obtained have been judged as good, as 1) the sentences provide an understandable description of trends in the data in everyday natural language, and 2) the relevance heuristic adequately selects relevant aspects from the set of sentences.

The used methodology is flexible and has the potential to generate a wealth of sentences. Not only can other statistical indicators be used to diversify the possible descriptions of an interval, e.g. adding ‘early’ or ‘late’, but additionally many different comparisons can be made, e.g. between days, Mondays, weekdays, weeks, months, years, etc.

ACKNOWLEDGMENT

This work has been partially funded by the Foundation for the Advancement of Soft Computing (Asturias, Spain) and the Spanish government (CICYT) under grant: TIN2008-06890-C02-01. The authors would like to thank HC Energía providing the expertise and data needed for the successful completion of the project.

REFERENCES

[1] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds., *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996.

[2] M. Dunham, Ed., *Data Mining Introductory and Advanced Topics*. Prentice Hall, 2003.

[3] K. Cios, W. Pedrycz, R. Swiniarski, and L. Kurgan, Eds., *Data Mining: A Knowledge Discovery Approach*. Springer, 2007.

[4] G. Raschia and N. Mouaddib, “Saintetiq: a fuzzy set-based approach to database summarization,” *Fuzzy Sets and Systems*, vol. 129, no. 2, pp. 137–162, 2002.

[5] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[6] ——, “Fuzzy logic = computing with words,” *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 103–111, 1996.

[7] R. R. Yager and L. A. Zadeh, Eds., *An Introduction to Fuzzy Logic Applications in Intelligent Systems*. Norwell, MA, USA: Kluwer Academic Publishers, 1992.

[8] J. Kacprzyk and S. Zadrozny, “Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools,” *Information Sciences*, vol. 173, no. 4, pp. 281–304, 2005.

[9] S. Zadrożny, J. Kacprzyk, and M. Gola, “Towards human friendly data mining: Linguistic data summaries and their protoforms,” *Lecture Notes in Computer Science, Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005*, vol. 3697/2005, pp. 697–702, 2005.

[10] J. Kacprzyk and R. Yager, “Linguistic summaries of data using fuzzy logic,” *International Journal of General Systems*, vol. 30, pp. 33–154, 2001.

[11] J. Kacprzyk, A. Wilbik, and S. Zadrożny, “Linguistic summarization of time series using a fuzzy quantifier driven aggregation,” *Fuzzy Sets Syst.*, vol. 159, no. 12, pp. 1485–1499, 2008.

[12] R. R. Yager, “Fuzzy summaries in database mining,” in *CAIA ’95: Proceedings of the 11th Conference on Artificial Intelligence for Applications*. Washington, DC, USA: IEEE Computer Society, 1995, p. 265.

[13] Hc energía. [Online]. Available: http://www.h-c.es/

[14] L. A. Zadeh, “The concept of a linguistic variable and its application to approximate reasoning - i,” *Information Sciences*, vol. 8, no. 3, pp. 199–259, 1975.

[15] ——, “The concept of a linguistic variable and its application to approximate reasoning - ii,” *Information Sciences*, vol. 8, no. 4, pp. 301–357, 1975.

[16] ——, “The concept of a linguistic variable and its application to approximate reasoning - iii,” *Information Sciences*, vol. 9, no. 1, pp. 43–80, 1975.

[17] ——, “Outline of a new approach to the analysis of complex systems and decision processes,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, pp. 28–44, 1973.

[18] J. Kacprzyk, A. Wilbik, and S. Zadrozny, “Using a genetic algorithm to derive a linguistic summary of trends in numerical time series,” *Proceedings of 2006 International Symposium on Evolving Fuzzy Systems*, no. September 7-9, pp. 137–142, 2006.