An extended, specificity based approach to linguistic summarization of time series

Janusz Kacprzyk, Anna Wilbik*
Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
{kacprzyk, wilbik}@ibspan.waw.pl

Abstract

We reformulate and extend our previous works (cf. Kacprzyk, Wilbik and Zadrożny [7] – [17]), mainly towards a more complex evaluation of results on the linguistic summarization of time series meant as the derivation of an linguistic quantifier driven aggregation of partial trends with respect to the dynamics of change, duration and variability. We use Zadeh's calculus of linguistically quantified propositions but, in addition to the basic criterion of a degree of truth (validity), we also use a degree of specificity to make it possible to account for a frequent case that though the degree of truth of a very general (not specific) summary is high, its usefulness may be low due to its low specificity. We show an application to the absolute performance type analysis of daily quotations of an investment fund.

Keywords: linguistic data summary, fuzzy logic, time series, linguistic quantifier, specificity, investment fund.

1 Introduction

A rapidly growing amount of data rapidly in virtually all fields implies serious difficulties to a human being whose cognitive capacity is limited. Moreover, natural language is the only fully natural means of articulation and communication for a human being, and it is not the case for the existing information technology. To bridge this gap, it would be very helpful to be able to use as much natural language as possible to capture the contents and meaning of sets of data. For this purpose we will employ linguistic summarization of data (bases).

A linguistic summary of a data (base) is meant as a concise, human-consistent description of a (numerical) data set expressed in (quasi)natural language, and was introduced by Yager [32] and then further developed, and presented in an implementable form, by Kacprzyk and Yager [18], and Kacprzyk, Yager and Zadrożny [19]. The contents of a database is summarized via linguistically quantified propositions, along the lines of Zadeh's calculus of linguistically quantified propositions [38].

Here we deal with *time series* which are omnipresent, and exemplified by sales data, quotations of shares, etc. over a certain period of time. Traditionally, the analysis of time series data is by using statistical methods, they provide powerful tools but are not human consistent enough because of lacking a natural language connection. Our approach makes an explicit use of natural language but is clearly not meant to replace classic statistical analyses, rather complementing them. It is a derivative and extension of the original approach of Kacprzyk, Wilbik and Zadrożny [7] – [17]), mainly towards a more complex evaluation of results.

The analysis of time series data involves different elements (cf. Batyrshin and Sheremetov [2, 3]) but we will concentrate, for lack of space, on the specifics of our approach.

* Supported by the Polish Ministry of Science and Higher Education under Grant No. NN516 4309 33

First, we need to identify the consecutive parts of time series within which the data exhibit some uniformity as to their variability. Clearly, some variability must here be neglected, under an assumed granularity. Here, these consecutive parts of a time series are called trends, and are described by straight line segments. That is, we perform first a piece wise linear approximation of a time series and present time series data as a sequence of trends. The (linguistic) summaries of time series refer to the (linguistic) summaries of (partial) trends as meant above. For the construction of such a piecewise linear approximation, we use a modified version of the Sklansky and Gonzalez algorithm (cf. [28]) though many other methods can be used – cf. Keogh et al. [22] – [24].

The next step in the derivation of linguistic summaries of time series is an aggregation of the (characteristic features of) consecutive trends over an entire time span (horizon) assumed. We follow the idea initiated by Yager [32] and then shown more profoundly and in an implementable way in Kacprzyk and Yager [18], and Kacprzyk, Yager and Zadrożny [19], that the most comprehensive and meaningful will be a linguistic quantifier driven aggregation resulting in linguistic summaries exemplified by *"Most trends are short"* or *"Most long trends are increasing"* which are easily derived and interpreted using Zadeh's fuzzy logic based calculus of linguistically quantified propositions. Basically, such summaries are interpreted in terms of the number or proportion of elements possessing a certain property related to those possessing some other property, a less restrictive one. The classic Zadeh's approach, the Sugeno and Choquet integrals, the OWA operators, etc. can be employed. A new quality, and an increased generality was obtained by using Zadeh's [39] protoforms as proposed by Kacprzyk and Zadrożny [20].

In this paper we will employ the classic Zadeh's fuzzy logic based calculus of linguistically quantified propositions as in the source papers by Kacprzyk, Wilbik and Zadrożny [7] – [17]). However, we will use different protoforms of linguistic time series summaries which are easier comprehensible by domain experts. Moreover, in addition to the degree of truth (validity) we use a degree of specificity as the second criterion to make it possible to account for a frequent case that though the degree of truth of a very general (not specific) summary is high, its usefulness may be low due to its low specificity.

We will illustrate our analysis on a linguistic summarization of daily quotations over an eight year period of an investment (mutual) fund. We will present in detail the characteristic features of trends derived under some reasonable granulations, variability, trend duration, etc.

It should be noted that the paper is in line with some other modern approaches to linguistic summarization of time series reported in the literature. First, from a slightly more general perspective, one should refer to the `SumTime` project coordinated by the University of Aberdeen, an EPSRC Funded Project for Generating Summaries of Time Series Data [1]. In this project English summary descriptions of a time-series data set are sought by using advanced time series and NLG (natural language generation) technologies [29]. However, the linguistic descriptions obtained do not reflect an inherent imprecision (fuzziness) as in our approach.

Now we will proceed to the description of the particular elements of our approach.

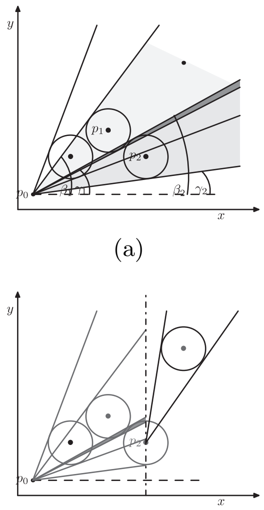
2 Preprocessing

A time series is a sequence of numerical data measured at uniformly spaced moments. We identify segments as linearly increasing, stable or decreasing functions, with a variable intensity, and therefore represent given time series data as piecewise linear functions. These are clearly partial trends as a global trend in a time series concerns the entire time span of the time series, and there also may be trends that concern parts of the entire time span, but more than a particular window

[1] cf. `www.csd.abdn.ac.uk/research/sumtime/`

taken into account while extracting partial segments. There are many algorithms for extracting piecewise linear segments of time series, that include the on-line (sliding window) algorithms, bottom-up or top-down strategy. A very good overview of those algorithms is presented in [23, 24, 22].

In our works we use a very simple on-line algorithm [8, 9], a modification of Sklansky and Gonzalez one [28]. For lack of space we will not present details of the algorithm employed. For illustration we will present its essence in Fig. 1.



(a)

(b)

Figure 1: An illustration of the algorithm for the uniform ε-approximation

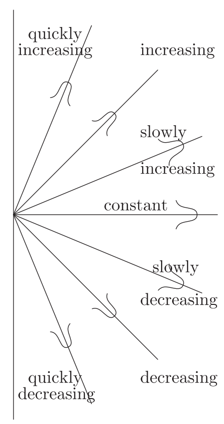


Figure 2: A visual representation of angle granules defining the dynamics of change

Basically, in our source approach (cf. Kacprzyk, Wilbik and Zadrożny [7] – [17]) we have considered the following three aspects of trends in time series: (1) dynamics of change, (2) duration, and (3) variability, and it should be noted that by trends we mean here global trends, concerning the entire time series (or some, probably large, part of it), not partial trends concerning a small time span (window) taken into account in the (partial) trend extraction phase via the used segmentation algorithm. These three characteristic features of trends are clearly the most straightforward and intuitively appealing ones as they concern those aspects of what happens with data over time that can easily be understood by domain experts. This has been clearly visible in our case while working with domain experts in the field of finance. These aspects will also be used in this paper.

Dynamics of change means the speed of change of the particular (consecutive) values of time series. It can be described by the slope of a line representing the trend, (cf. any angle η from the interval $\langle \gamma, \beta \rangle$ in Fig. 1(a)). Thus, to quantify the dynamics of change we may use the interval of possible angles $\eta \in \langle -90; 90 \rangle$. For practical reasons we use a fuzzy granulation to represent the values, for instance as in Fig. 2. For some methods to obtain such a granulation, cf. In Batyrshin et al. [1, 2]. We map a single value α (or the interval corresponding to the grey area in Fig. 1(b)) into a fuzzy set (linguistic label) best matching a given angle using some measure of a distance or similarity, cf. the book by Cross and Sudkamp [5].

Duration describes the length of a single trend, meant as a linguistic variable whose linguistic value (label) may be exemplified by a "long trend" defined as a fuzzy set with a properly defined membership function.

Variability refers to how "spread out" ("vertically", in the sense of values taken on) a group of data is. Previously and also here, we use a weighted average of values taken by some of the following measures used in statistics: (1) The range (maximum – minimum), (2) The interquartile range (IQR) calculated as the third quartile (the 75th percentile) minus the first quartile (the 25th percentile), (3) The variance, (4) The standard deviation, (5) The mean absolute deviation (MAD).

Similarly as in the case of dynamics of change, we find for a given value of variability obtained as above a best matching fuzzy set (linguistic label) using, e.g., some measure of a distance or similarity.Again, the measure of variability is treated as a linguistic variable and expressed using linguistic terms (labels) modeled by fuzzy sets defined by the user.

3 Linguistic data summaries

A linguistic summary is meant as a (usually short) natural language like sentence(s) subsuming the very essence of a set of data which is numeric and usually too large to be comprehended by the human being. (cf. Kacprzyk and Zadrożny [20], [21]).

In Yager's [32] approach (cf. rather Kacprzyk and Yager [18], and Kacprzyk, Yager and Zadrożny [19] for a more realistic and implementable version) the following perspective for linguistic data summaries is assumed:
– $Y = \{y_1, \ldots, y_n\}$ is a set of objects (records) in a database, e.g., the set of workers;
– $A = \{A_1, \ldots, A_m\}$ is a set of attributes characterizing objects from Y, e.g., salary, age, etc. in a database of workers, and $A_j(y_i)$ denotes a value of attribute A_j for object y_i.

A linguistic summary of a data set comprises:
– a summarizer P, i.e. an attribute together with a linguistic value (fuzzy predicate) defined on the domain of attribute A_j (e.g. "low" for attribute "salary");
– a quantity in agreement Q, i.e. a linguistic quantifier (e.g. most);
– truth (validity) $\mathcal{T}$ of the summary, i.e. a number from the interval $[0, 1]$ assessing the truth (validity) of the summary (e.g. 0.7); usually, only summaries with a high value of $\mathcal{T}$ are interesting;
– and, optionally, a qualifier R, i.e. another attribute together with a linguistic value (fuzzy predicate) defined on the domain of attribute A_k determining a (fuzzy subset) of Y (e.g. "young" for attribute "age"),
and can be exemplified by

$$\mathcal{T}(\text{most of employees earn low salary}) = 0.7 \quad (1)$$

or, in an extended form including a qualifier (e.g. young), by

$$\mathcal{T}(\text{most of young employees earn low salary}) = 0.9 \quad (2)$$

Thus, in this approach the core of a linguistic summary is a *linguistically quantified proposition* in the sense of Zadeh [38] which, for (1) and (2), respectively, may be written as:

$$Qy\text{'s are } P \quad (3)$$

$$QRy\text{'s are } P \quad (4)$$

Then, $\mathcal{T} \in [0, 1]$, i.e., the truth (validity) of a linguistic summary, directly corresponds to the truth value of (3) or (4). This may be calculated by using either the original Zadeh's fuzzy logic based calculus of linguistically quantified propositions (cf. [38]) yielding, respectively:

$$\mathcal{T}(Qy\text{'s are } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \quad (5)$$

$$\mathcal{T}(QRy\text{'s are } P) = \;=\; \mu_Q\left(\frac{\sum_{i=1}^{n}(\mu_R(y_i) \wedge \mu_P(y_i))}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \quad (6)$$

where $\wedge$ is the minimum operation (more generally it can be another appropriate operation, notably a t-norm), and Q is a fuzzy set representing the linguistic quantifier in the sense of Zadeh [38], i.e. $\mu_Q : [0, 1] \longrightarrow [0, 1]$, $\mu_Q(x) \in [0, 1]$.

We consider basically *regular non-decreasing monotone* quantifiers such that:

$$\mu(0) = 0, \quad \mu(1) = 1 \quad (7)$$

$$x_1 \leq x_2 \Rightarrow \mu_Q(x_1) \leq \mu_Q(x_2) \quad (8)$$

which can be exemplified by "most" given as:

$$\mu_Q(x) = \begin{cases} 1 & \text{for } x \geq 0.8 \\ 2x - 0.6 & \text{for } 0.3 < x < 0.8 \\ 0 & \text{for } x \leq 0.3 \end{cases} \quad (9)$$

Other methods of calculating $\mathcal{T}$ can be used, notably those based on the OWA (ordered weighted averaging) operators (cf. Yager [33, 34], Yager and Kacprzyk [36]), and the Sugeno and Choquet integrals (cf. Bosc and Lietard [4] or Grabisch [6]).

4 Protoforms of linguistic trend summaries

Kacprzyk and Zadrożny [20] showed that Zadeh's [39] concept of a protoform is convenient for dealing with linguistic summaries. This approach is also employed here.

Basically, a protoform is defined as a more or less abstract prototype (template) of a linguistically quantified proposition. In our context of time series summaries, the use of protoforms was proposed by Kacprzyk, Wilbik and Zadrożny[7] – [17]. In this paper we use different types of protoforms of time series summaries which are more clear to the practitioners:
– for a short short form:

$$\text{Among all segments, } Q \text{ are } P \tag{10}$$

– for an extended form:

$$\text{Among all } R \text{ segments, } Q \text{ are } P \tag{11}$$

A detailed exposition and analysis of these kinds of protoforms, in a static and dynamic (for the time series summaries) will be presented in a separate paper.

5 Quality measures

In Kacprzyk, Wilbik and Zadrożny's [7] – [17] works, in which the new approach to the linguistic summarization of time series has been proposed, the basic quality criterion of the truth value of a linguistic summary was employed. This may be sufficient in many practical cases, and the numerical simplicity is an additional relevant feature. However, similarly as for the traditional linguistic summaries of static data sets (data bases) when the degree of truth was proposed in the source Yager's [32] paper and then used in practically all his next papers, some other quality measures has been later proposed, notably by Kacprzyk and Yager [18] and Kacprzyk, Yager and Zadrożny [19].

The following basic quality measures of linguistic summaries can be distinguished: (1) a truth value, (2) a degree of imprecision, (3) a degree of (non)specificity, (4) a degree of fuzziness, (5) a degree of covering (support), (6) a degree of appropriateness, and (7) the length of the summary.

In this paper we will use, first, the traditional truth value, and then – as a second criterion – the degree of specificity.

The truth value (a degree of truth), introduced by Yager in [32], is the basic criterion describing the degree of truth (from $[0,1]$) to which a linguistically quantified proposition equated with a linguistic summary is true. It is calculated for the simple and extended forms of the summary, as, respectively:

$$\begin{aligned}&\mathcal{T}(\text{Among all } Y,\ Q \text{ are } P)\\ &\qquad = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right)\end{aligned} \tag{12}$$

$$\begin{aligned}&\mathcal{T}(\text{Among all } RY,\ Q \text{ are } P) =\\ &\quad = \ \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_R(y_i)\wedge\mu_P(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right)\end{aligned} \tag{13}$$

There are a few possibilities of calculating the nonspecificity of a fuzzy set. One is the standard nonspecificity measure from fuzzy sets theory, using the so called Hartley function (cf. Klir and Wierman [25] or Klir and Yuan [26]). For a finite, nonempty (crisp) set, A, we measure this amount using a function from the class of functions

$$U(A) = c\log_b |A|, \tag{14}$$

where $|A|$ dentoes the cardinality of A, b and c are positive constants, $b, c > 1$ (usually, $b = 2$ and $c = 1$). Then

$$U(A) = \log_2 |A|. \tag{15}$$

This function is applicable to finite sets only. However, it may be appropriately modified for infinite sets of $\mathbb{R}$:

$$U(A) = \log[1 + \mu(A)], \tag{16}$$

where $\mu(A)$ is the measure if A defined by the Lebesque integral of the characteristic function of A. When $A = [a, b]$, than $\mu(A) = b - a$ and $U([a, b]) = \log[1 + b - a]$.

For any nonempty fuzzy set A defined on a finite universal set X, the function $U(A)$ has the form

$$U(A) = \frac{1}{h(A)}\int_0^{h(A)} \log_2 |A^\alpha| d\alpha, \tag{17}$$

where $|A^\alpha|$ denotes the cardinality of the α-cut of A and $h(A)$ is the height of A.

When a nonempty fuzzy set is defined in the set of reals $\mathbb{R}$ and the α-cuts are infinitive sets (e.g., intervals of real numbers), we have to calculate $U(A)$ using:

$$U(A) = \frac{1}{h(A)} \int_0^{h(A)} \log[1 + \mu(A^\alpha)] d\alpha, \quad (18)$$

Other solutions are also possible (cf. Yager, Ford and Canas [35]).

For convenience, the value of specificity is normalized.

In most applications, both the fuzzy predicates P and R are assumed to be of a rather simplified, atomic form referring to just one attribute. They can be extended to cover more sophisticated summaries involving some confluence of various attribute values as, e.g, "slowly decreasing and short" trends. To combine more then one attribute values we will use t-norms (the minimum or product) for conjunction and a corresponding s-norm (the maximum or probabilistic sum) for disjunction.

Then the degree of specificity of "Among all Y, Q are P" is:

$$\begin{aligned} & d_s(\text{"Among all } Y, Q \text{ are } P\text{"}) \\ & \quad = \quad 1 - (U(P) \wedge U(Q)) \end{aligned} \quad (19)$$

and the degree of specificity of "Among all RY, Q are P" is:

$$\begin{aligned} & d_s(\text{"Among all } RY, Q \text{ are } P\text{"}) \\ & \quad = \quad 1 - (U(P) \wedge U(R) \wedge U(Q)) \end{aligned} \quad (20)$$

where $U(P)$ is the degree of nonspecificity of the summarizer P, given by (18), $U(R)$ is the degree of nonspecificity of the qualifier R, $U(Q)$ is the degree of nonspecificity of the quantifier Q, and $\wedge$ is a t-norm (minimum or product).

6 Numerical experiments

The method proposed in this paper was tested on data on quotations of an investment (mutual) fund that invests at most 50% of assets in shares listed at the Warsaw Stock Exchange. Data shown in Figure 3 were collected from April 1998 until July 2007 with the value of one share equal to PLN 10.00 in the beginning of the period to PLN 55.27 at the end of the time span considered (PLN stands for the Polish Zloty). The minimal value recorded was PLN 6.88 while the maximal one during this period was PLN 57.85. The biggest daily increase was equal to PLN 1.27, while the biggest daily decrease was equal to PLN 2.41.

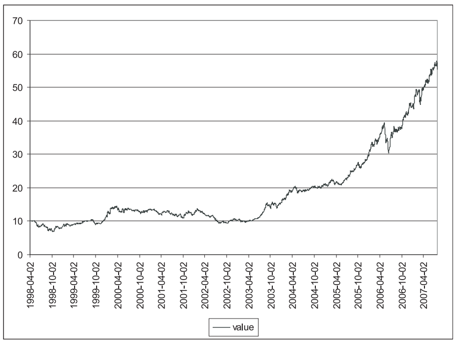


Figure 3: Daily quotations of an investment fund in question

Using the Sklansky and Gonzalez algorithm and $\varepsilon = 0.25$ we obtained 326 extracted (partial) trends. The shortest trend took 2 days, while the longest 71 days. The histogram of the duration of trends is presented in Fig. 4. Figure 5 shows the histogram of angles which characterize the dynamics of change. The histogram of the variability of trends (in %), assumed to be – for simplicity – the interquartile range only, is presented in Fig. 6.

There may be various ways of using linguistic summaries of trends for making decisions concerning various aspects, mainly investment decisions in our context. In this paper we will show an absolute performance type analysis, that is we will just deal with the values (price quotations) of shares of the investment fund in question. We will not deal with relations between these values and some appropriate benchmarks exemplified by daily values of some stock market index, or an appropriate mix of indexes, daily percentual change (absolute or related to the daily percentual change of some benchmarks), etc. as is often done by professionals.

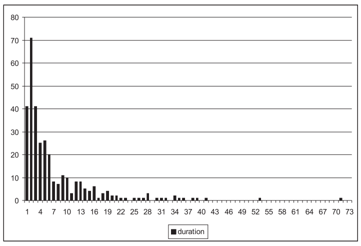


Figure 4: Histogram of duration of trends (in the number of days)

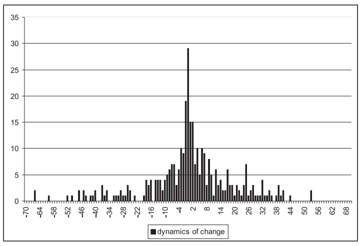


Figure 5: Histogram of angles (in degrees) characterizing the dynamic of change

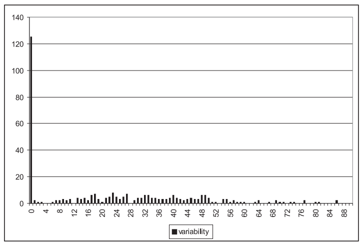


Figure 6: Histogram of the variability (the interquartile range) of trends

Some interesting summaries obtained by using the method proposed, employing the classic Zadehs calculus of linguistically quantified propositions [with properly defined elements like the fuzzy linguistic quantifiers, like (9) for "most"] and the Hartley function for the nonspecificity, and for different granulations of the dynamics of change, duration and variability, are:

– for 3 labels for the dynamics of change (*decreasing, constant, increasing*), the duration (*short, medium length, long*) and the variability (*low, medium, high*) are shown in Tab. 1.

Table 1:

linguistic summary	$\mathcal{T}$	d_s	α
Among all trends, most are short	0.713	0.902	0.760
Among all trends, most are constant	0.632	0.786	0.671
Among all trends, most are of a low variability	0.703	0.586	0.674
Among all short trends, most are of a low variability	0.878	0.902	0.884
Among all trends of a low variability, most are short	0.890	0.902	0.893
Among all increasing trends, most are of a low variability	0.887	0.586	0.811
Among all medium trends, at least around a half is of medium variability	1.000	0.820	0.955
Among all trends of a high variability, at least around a half is increasing	0.916	0.586	0.834
Among all decreasing trends, almost all are short	1.000	0.902	0.976

– for 5 labels for the dynamics of change (*quickly decreasing, decreasing, constant, increasing, quickly increasing*), the duration (*very short, short, medium length, long, very long*) and the variability (*very low, low, medium, high, very high*) are in Tab. 2.

Table 2:

linguistic summary	$\mathcal{T}$	d_s	α
Among all trends, most are constant	0.632	0.786	0.671
Among all trends, at least around a half is very short	1.000	0.934	0.984
Among all trends, at least around a half is of a very low variability	1.000	0.772	0.943
Among all very short trends, most are of a very low variability	0.935	0.934	0.935
Among all trends of a very low variability, most are very short	1.000	0.934	0.984
Among all constant trends, at least around a half are short	0.748	0.887	0.783
Among all trends of a very high variability, at least around a half are quickly increasing	0.830	0.772	0.816
Among all trends of medium length, almost all are constant	1.000	0.820	0.955
Among all quickly decreasing trends, almost all are very short	1.000	0.934	0.984
Among all long trends, much more than a half is of a low variability	0.552	0.707	0.591

– for 7 labels for the dynamics of change (*quickly decreasing, decreasing, slowly decreasing, constant, slowly increasing, increasing, quickly increasing*), the duration (*very short, short, rather short, medium length, rather long, long, very long*) and the variability (*very low, low, rather low, medium, rather high, high, very high*) are in Tab. 3.

Table 3:

linguistic summary	$\mathcal{T}$	d_s	α
Among all trends, almost none are of rather high variability	1.000	0.879	0.970
Among all trends, at least around a half are very short	1.000	0.934	0.984
Among all trends, at least around a half are constant	1.000	0.867	0.967
Among all trends, at least around a half are of very low variability	0.914	0.806	0.887
Among all trends, at most around one third is slowly increasing	1.000	0.909	0.977
Among all very short trends, most are of a very low variability	0.905	0.934	0.912
Among all trends of a very low variability, most are very short	1.000	0.934	0.984
Among all slowly decreasing trends, most are very short	0.971	0.934	0.962
Among all constant trends, at least around a half is short	0.683	0.887	0.734
Among all trends of medium length, almost all are constant	1.000	0.918	0.980
Among all quickly decreasing trends, almost all are very short	1.000	0.934	0.984
Among all long trends, much more than a half is of rather low variability	0.734	0.842	0.761

7 Concluding remarks

We reformulated and extended our previous works on linguistic summarization of time series by adding to the basic quality criterion of a degree of truth (validity), a degree of specificity. This made it possible to make a more reasonable choice between the summaries obtained as, frequently, though the degree of truth of a very general (not specific) summary may be high, its usefulness may be low. The results obtained on the analysis of the absolute performance of daily quotations of an investment fund seem very promising.

References

[1] I. Batyrshin, On granular derivatives and the solution of a granular initial value problem, International Journal Applied Mathematics and Computer Science 12 (3) (2002) 403–410.

[2] I. Batyrshin, L. Sheremetov, Perception based functions in qualitative forecasting. I. Batyrshin et al. (Eds.), Perception-based Data Mining and Decision Making in Economics and Finance, Springer-Verlag (2006) 119–134.

[3] I. Batyrshin, L. Sheremetov, Towards perception based time series data mining. M. Nikravesh et al.(Eds.), Forging New Frontiers. Fuzzy Pioneers I, Springer-Verlag (2007) 217–230.

[4] P. Bosc, L. Lietard, O. Pivert, Quantified statements and database fuzzy queries. P. Bosc, J. Kacprzyk (Eds.), Fuzziness in Database Management Systems, Springer-Verlag (1995).

[5] V. Cross, T. Sudkamp, Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications, Springer-Verlag (2002).

[6] M. Grabisch, Fuzzy integral as a flexible and interpretable tool of aggregation. B. Bouchon-Meunier (Ed.), Aggregation and Fusion of Imperfect Information, Physica–Verlag (1998) 51–72.

[7] J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic summarization of trends: a fuzzy logic based approach. Proceedings of the 11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems, (2006) 2166–2172.

[8] J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic summaries of time series via a quantifier based aggregation using the Sugeno integral. Proceedings of 2006 IEEE World Congress on Computational Intelligence, IEEE Press, (2006) 3610–3616.

[9] J. Kacprzyk, A. Wilbik, S. Zadrożny,

On some types of linguistic summaries of time series. Proceedings of the 3rd International IEEE Conference Intelligent Systems, IEEE Press (2006) 373–378.

[10] J. Kacprzyk, A. Wilbik, S. Zadrożny, A linguistic quantifier based aggregation for a human consistent summarization of time series. J. Lawry et al.(Eds.), Soft Methods for Integrated Uncertainty Modelling. Springer-Verlag (2006) 186–190.

[11] J. Kacprzyk, A. Wilbik, S. Zadrożny, Capturing the essence of a dynamic behavior of sequences of numerical data using elements of a quasi-natural language. Proceedings of the 2006 IEEE International Conference on Systems, Man, and Cybernetics, IEEE Press (2006) 3365–3370.

[12] J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic summarization of time series by using the Choquet integral. P. Melin et al.(Eds.), Foundations of Fuzzy Logic and Soft Computing IFSA 2007, LNAI 4529, Springer-Verlag (2007) 284-294.

[13] J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic summarization of time series under different granulation of describing features. M. Kryszkiewicz et al.(Eds.), Rough Sets and Intelligent Systems Paradigms - RSEISP 2007, LNAI 4585, Springer-Verlag(2007) 230-240.

[14] J. Kacprzyk, A. Wilbik, S. Zadrożny, Analysis of time series via their linguistic summarization: the use of the Sugeno integral. L. de Macedo Mourelle et al.(Eds.), Proceedings of the 7th International Conference on Intelligent Systems Design and Applications ISDA 2007, IEEE Press (2007) 262-267.

[15] J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic Summaries of Time Series via an OWA Operator Based Aggregation of Partial Trends. Proceedings of the FUZZ-IEEE 2007 IEEE International Conference on Fuzzy Systems, IEEE Press (2007) 467-472.

[16] J. Kacprzyk, A. Wilbik, S. Zadrożny, Mining time series data via linguistic summaries of trends by using a modified Sugeno integral based aggregation. Proceedings of IEEE Symposium on Computational Intelligence and Data Mining, IEEE Press (2007) 742-749.

[17] J. Kacprzyk, A. Wilbik, S. Zadrożny, On Linguistic Summaries of Time Series via a Quantifier Based Aggregation Using the Sugeno Integral. O. Castillo et al.(Eds.), Hybrid Intelligent Systems Analysis and Design, Springer-Verlag (2007) 421-439.

[18] J. Kacprzyk, R. R. Yager, Linguistic summaries of data using fuzzy logic, International Journal of General Systems 30 (2001) 33–154.

[19] J. Kacprzyk, R. R. Yager, S. Zadrożny, A fuzzy logic based approach to linguistic summaries of databases, International Journal of Applied Mathematics and Computer Science 10 (2000) 813–834.

[20] J. Kacprzyk, S. Zadrożny, Linguistic database summaries and their protoforms: toward natural language based knowledge discovery tools, Information Sciences 173 (2005) 281–304.

[21] J. Kacprzyk, S. Zadrożny, Fuzzy linguistic data summaries as a human consistent, user adaptable solution to data mining. B. Gabrys et al.(Eds.), Do Smart Adaptive Systems Exist?, Springer, (2005) 321–339.

[22] E. Keogh, M. Pazzani, An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. Procedings of the 4th International Conference on Knowledge Discovery and Data Mining (1998) 239-241.

[23] E. Keogh, S. Chu, D. Hart, and M. Pazzani, An Online Algorithm for Segmenting Time Series, Proceedings of IEEE International Conference on Data Mining (2001) 289–296.

[24] E. Keogh, S. Chu, D. Hart, and M. Pazzani, Segmenting Time Series: A Survey and Novel Approach. M. Last et al.(Eds.), Data Mining in Time Series Databases, World Scientific Publishing (2004).

[25] G. J. Klir, M. J. Wierman, Uncertainty-Based Information, Elements of Generalized Information Theory, Physica-Verlag, 1999.

[26] G. J. Klir, B. Yuan, Fuzzy Stes and Fuzzy Logic, Theory and Applications, Prentice Hall, 1995.

[27] M. Last, Y. Klein, A. Kandel, Knowledge discovery in time series databases. IEEE Trans. on Syst., Man and Cybern. SMC, Part B, 31 (2001) 160–169.

[28] J. Sklansky, V. Gonzalez, Fast polygonal approximation of digitized curves, Pattern Recognition 12 (5) (1980) 327–331.

[29] S. Sripada, E. Reiter, I. Davy. SumTime-Mousam: Configurable Marine Weather Forecast Generator. Expert Update 6(3) (2003) 4–10.

[30] G. Stockman, L. Kanal, M.C. Kyle, Structural pattern recognition of carotid pulse waves using a general waveform parsing system. Comm. of ACM 19 (1976) 688–695.

[31] StockReporter: cf. `http://www.ics.mq.edu.au/~ltgdemo/StockReporter/about.html`.

[32] R.R. Yager, A new approach to the summarization of data, Information Sciences 28 (1982) 69–86.

[33] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, IEEE Transactions on Systems, Man and Cybernetics SMC-18 (1988) 183–190.

[34] R.R. Yager, Quantifier guided aggregation using OWA operators, International Journal of Intelligent Systems 11 (1996) 49–73.

[35] R.R. Yager, K.M. Ford, A.J. Canas, An Approach to the linguistic summarization of data, B. Bouchon-Meunier et al.(Eds.) Uncertainty in Knowledge Bases: Proc. of the 3rd IPMU'90, Springer (1991) 456-468.

[36] R.R. Yager, J. Kacprzyk, The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer, Boston (1997).

[37] J. Yu, E. Reiter, J. Hunter, C. Mellish, Choosing the content of textual summaries of large time-series data sets. Natural Language Engineering 2007; 13(1): 25-49.

[38] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, Computers and Mathematics with Applications 9 (1983) 149–184.

[39] L.A. Zadeh, A prototype-centered approach to adding deduction capabilities tosearch engines – the concept of a protoform. Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2002), (2002) 523–525.