123/2009

Raport Badawczy
Research Report

RB/54/2009

Linguistic summaries of time series: on some additional data independent quality criteria

J. Kacprzyk, A. Wilbik

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Janusz Kacprzyk

Warszawa 2009

Linguistic summaries of time series: on some additional data independent quality criteria

Janusz Kacprzyk   Anna Wilbik

Abstract

We further extend our approach on the linguistic summarization of time series (cf. Kacprzyk, Wilbik and Zadrożny [19, 18, 20, 17, 16, 23, 24, 21, 22, 25, 26]) in which an approach based on a calculus of linguistically quantified propositions is employed, and the essence of the problem is equated with a linguistic quantifier driven aggregation of partial scores (trends). Basically, we present here some reformulation and extension of our works mainly by including a more complex evaluation of the linguistic summaries obtained. In addition to the basic criterion of a degree of truth (validity), we also use he as the additional criteria a degree of imprecision, specificity, fuzziness and focus. However, for simplicity and tractability, we use in the first shot the degrees of truth (validity) and focus, which usually reduce the space of possible linguistic summaries to a considerable extent, and then – for a usually much smaller set of linguistic summaries obtained – we use the remaining 3 degrees of imprecision, specificity and fuzziness for making a final choice of appropriate linguistic summaries. We show an application to the absolute performance type analysis of daily quotations of an investment fund.

1 Introduction

Financial data analysis is one of the most important application areas of advanced data mining and knowledge discovery tools and techniques. For instance, in a report presented by G. Piatetsky-Shapiro's KDNuggets (http://www.kdnuggets.com) on top data mining applications in 2008, the first two positions are, in the sense of yearly increase:

- *Investment/Stocks*, up from 3% of respondents in 2007 to 14% of respondents in 2008% (350% increase),
- *Finance*, up form 7.2% in 2007 to 16.8% in 2008 (108% increase).

This general trend will presumably continue over the next years, maybe decades, in view of a world wide financial and economic that are expected to continue well after 2009.

This paper is a continuation of our previous works (cf. Kacprzyk, Wilbik, Zadrożny [19, 20, 23, 22, 25, 27] or Kacprzyk, Wilbik [12, 14, 13]) which deal with the problem of how to effectively and efficiently support a human decision maker in making decisions concerning investments. We deal mainly with in investment (mutual) funds. Clearly, decision makers are here concerned with possible future gains/losses, and their decisions is related to what might happen in the future.

However, our aim is not the forecasting of the future daily prices, which could have been eventually used directly for a purchasing decision. Instead, in our works, we follow a decision support paradigm (Fig. 1), that is we try to provide the decision maker with some information that can be useful for his/her decision on whether and how many units of funds to purchase. We do not intend to replace the human decision maker.

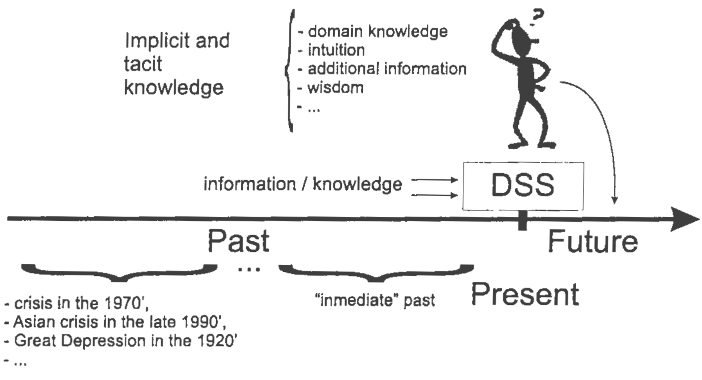

Figure 1: Decision support paradigm

This problem is very complex. First of all, there may be two general approaches. The first one, which may seem to be the most natural is to provide means to derive a price forecast for an investment unit so that the decision maker could "automatically" purchase what has been forecast, and as much as he/she could afford. Unfortunately, the success in such a straightforward approach has been much less than expected. Basically, statistical methods employed usually for this purpose are primitive in the sense that they just extrapolate the past and do not use domain knowledge, intuition, some inside information, etc. A natural solution may be to try to just support the human decision maker in making those investment decisions by providing him/her with some additional useful information, and not getting involved in the actual investment decision making.

Various philosophies in this respect are possible. Basically, from our perspective, the following one will be followed. In all investment decisions the future is what really counts, and the past is irrelevant. But, the past is what we know, and the future is (completely) unknown. Behavior of the human being is to a large extend driven by his/her (already known) past experience. We usually assume that what happened in the past will also happen (to some, maybe large extent) in the future. This is basically, by the way, the very underlying assumption behind the statistical methods too!

This clearly indicates that the past can be employed to help the human decision maker find a good solution. We present here a method to subsume the past, the past performance of an investment (mutual) fund, by presenting results in a vary human consistent way, using natural language statements.

We will apply our method to mutual funds quotations, as those time series are easily available, and almost everyone can invest its money in a mutual fund. However if one looks at an information leaflet, one may always notice a disclaimer stating that "Past performance is no indication of future returns" which is true. However, on the other hand, in a well known posting "Past Performance Does Not Predict Future Performance" [3], they state something that may look strange in this context, namely: "...according to an Investment Company Institute study, about 75% of all mutual fund investors mistakenly use short-term past performance as their primary reason for buying a specific fund". But, in an equally well known posting "Past performance is not everything" [4], they state: "...disclaimers apart, as a practice investors continue to make investments based on a schemes past performance. To make matters worse, fund houses are only too pleased to toe the line by actively advertising the past performance of their schemes leading investors to conclude that it is the single-most important parameter (if not the most important one) to be considered while investing in a mutual fund scheme".

As strange as this apparently is, we may ask ourselves why it is so. Again, in a well known posting "New Year's Eve:Past performance is no indication of future return" [2], they say "...if there is no correlation between past performance and future return, why are we so drawn to looking at charts and looking at past performance? I believe it is because it is in our nature as human beings ... because we don't know what the future holds, we look toward the past ...".

And, continuing along this line of reasoning, we can find many other examples of similar statements supporting our position. For instance, in [35], author says: "...Does this mean you should ignore past performance data in selecting a mutual fund? No. But it does mean that you should be wary of how you use that information ... While some research has shown that consistently good performers continue to do well at a better rate than marginal performers, it also has shown a much stronger predictive value for consistently bad performers ... *Lousy performance in the past is indicative of lousy performance in the future...*". And, further: in [7], we have: "... there is an important role that past performance can play in helping you to make your fund selections. While you should disregard a single aggregate number showing a fund's past long-term return, you can learn a great deal by studying the *nature of its past returns*. Above all, look for consistency.". In [37], we find: "While past performance does not necessarily predict future returns, it can tell you how volatile a fund has been". In the popular "A 10-step guide to evaluating mutual funds" [1], they say in the last, tenth, advise: "Evaluate the funds performance. Every fund is benchmarked against an index like the BSE Sensex, Nifty, BSE 200 or the CNX 500 to cite a few names. Investors should compare fund performance over varying time frames vis-a-vis both the benchmark index and peers. Carefully evaluate the funds performance across market cycles particularly the downturns".

Therefore we think, that linguistic summaries may be easily understood by the humans and present them briefly the performance of the mutual fund, and this knowledge may be later incorporated while making up decisions.

Here we extend our previous works on linguistic summarization of time series (cf. Kacprzyk, Wilbik, Zadrożny [19, 20, 23, 22, 25, 27] or Kacprzyk, Wilbik [12, 14, 13]), mainly towards a more complex evaluation of results. Generally the basic criterion for evaluation linguistic summaries is a degree of truth (used by us at first in our papers [19, 20, 22, 27]) as it was originally proposed in the static context by Yager [40]. However later Kacprzyk and Yager [28] and Kacprzyk, Yager and Zadrożny [29, 30] and Kacprzyk and Zadrożny [11, 10] introduced some additional quality criteria, notably a degree of specificity, fuzziness, and imprecision.

In this paper we will discuss the degree of imprecision, as well as two others measures similar in spirit, namely degree of specificity and degree of fuzziness.

2 Linguistic data summaries

Under the term linguistic summary of data (base) we understand a (usually short) sentence (or a few sentences) that captures the very essence of the set of data, that is numeric, large, and because of its size is not comprehensible for human being.

In Yager's basic approach [40], and later papers on this topic, as well as here the following notation is used:

- $Y = \{y_1, y_2, \ldots, y_n\}$ is the set of objects (records) in the database D, e.g., a set of employees;
- $A = \{A_1, A_2, \ldots, A_m\}$ is the set of attributes (features) characterizing objects from Y, e.g., a salary, age in the set of employees.

A linguistic summary includes:

- 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);
- 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*).

Thus, a linguistic summary may be exemplified by

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

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

$$\mathcal{T}(\textit{most} \text{ of } \textit{young} \text{ employees earn } \textit{low} \text{ salary}) = 0.82 \tag{2}$$

Thus, basically the core of a linguistic summary is a linguistically quantified proposition in the sense of Zadeh [46] which for (1) may be written as

$$Qy's \text{ are } P \tag{3}$$

and for (2) may be written as

$$QRy's \text{ are } P \tag{4}$$

Then the truth (validity), $\mathcal{T}$, of a linguistic summary directly corresponds to the truth value of (3) and (4). This may be calculated using either original Zadeh's calculus of quantified propositions (cf. [46]) or other interpretations of linguistic quantifiers. In the former case the truth values of (3) and (4) are calculated, respectively, as

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

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

where $\wedge$ is the minimum operation (more generally it can be another appropriate operator, notably a t-norm), and Q is a fuzzy set representing the linguistic quantifier in the sense of Zadeh [46], i.e. regular, nondecreasing and monotone:

(a) $\mu_Q(0) = 0$,

(b) $\mu_Q(1) = 1$, and

(c) if $x > y$, then $\mu_Q(x) \geq \mu_Q(y)$;

It may be exemplified by *most* given by

$$\mu_Q(x) = \begin{cases} 1 & \text{for } x \geq 0.8 \\ 2x - 0.6 & \text{for } 0.3 < x < 0.8 \\ 0 & \text{for } x \leq 0.3 \end{cases} \tag{7}$$

Other methods of calculating $\mathcal{T}$ can be used here, notably those based on OWA (ordered weighted averaging) operators (cf. Yager [41, 43] and Yager and Kacprzyk [45]), and the Sugeno and Choquet integrals (cf. Bosc and Lietard [8] or Grabisch [9]).

3 Linguistic summaries of trends

In our first approach we summarize the trends (segments) extracted from time series. Therefore as the first step we need to extract the segments. We assume that segment is represented by a fragment of straight line, because such segments are easy for interpretation.

There are many algorithms for the piecewise linear segmentation of time series data, including e.g. on-line (sliding window) algorithms, bottom-up or top-down strategies (cf. Keogh [31, 32]). In our works [19, 18, 17, 16, 23, 24, 22, 25, 26] we used a simple on-line algorithm, a modification of the Sklansky and Gonzalez one [38].

We consider the following three features of (global) trends in time series:

1. dynamics of change,
2. duration, and
3. variability.

By *dynamics of change* we understand the speed of change of the consecutive values of time series. It may be described by the slope of a line representing the trend, represented by a linguistic variable.

Duration is the length of a single trend, and is also represented by a linguistic variable.

Variability describes how "spread out" a group of data is. We compute it as a weighted average of values taken by some measures used in statistics: (1) the range, (2) the interquartile range (IQR), (3) the variance, (4) the standard deviation, and (5) the mean absolute deviation (MAD). This is also treated as a linguistic variable.

For practical reasons for all we use a fuzzy granulation (cf. Bathyrshin at al. [5, 6]) to represent the values by a small set of linguistic labels as, e.g.: quickly increasing, increasing, slowly increasing, constant, slowly decreasing, decreasing, quickly decreasing. These values are equated with fuzzy sets.

For clarity and convenience we employ Zadeh's [47] protoforms for dealing with linguistic summaries [11]. A protoform is defined as a more or less abstract prototype (template) of a linguistically quantified proposition. We have two types of protoforms of linguistic summaries of trends:

- a short form:

$$\text{Among all segments, } Q \text{ are } P \tag{8}$$

  e.g.: "Among all segments, *most* are *slowly increasing*".

- an extended form:

$$\text{Among all } R \text{ segments, } Q \text{ are } P \tag{9}$$

  e.g.: "Among all *short* segments, *most* are *slowly increasing*".

The protoforms are very convenient for various reasons, notably: they make it possible to devise general tools and techniques for dealing with a variety of statements concerning different domains and problems, and their form is often easily comprehensible to domain specialists.

In static context Kacprzyk and Yager [28], Kacprzyk, Yager and Zadrożny [29, 30], and Kacprzyk and Zadrożny [11, 10] proposed several additional quality criteria, except from the basic one, the truth value. Those was, among others, degree of imprecision. We will discuss it here, as well as two others measures similar in spirit, namely degree of specificity and degree of fuzziness.

Generating the set of summaries requires checking many possible summaries and lots of time. However we follow a simplified approach it that we use a two-level procedure. First we reduce the search space of possible linguistic summaries, for this purpose we use the truth value and the degree of focus. And then we use the remaining degrees of imprecision, specificity and fuzziness. So this heuristic method does not guarantee the optimality, our experience however suggests that it makes possible to generate good summaries in computationally reasonable time.

3.1 Truth value

The truth value (a degree of truth or validity), introduced by Yager in [40], 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.

Using Zadeh's calculus of linguistically quantified propositions [46] it is calculated in dynamic context using the same formulas as in the static case. Thus, the truth value is calculated for the simple and extended form as, respectively:

$$\mathcal{T}(\text{Among all } y\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \tag{10}$$

$$\mathcal{T}(\text{Among all } Ry\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_R(y_i)\wedge\mu_P(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \tag{11}$$

where $\wedge$ is the minimum operation (more generally it can be another appropriate operator, notably a t-norm). In Kacprzyk, Wilbik and Zadrożny [24] results obtained by using different t-norms were compared. Various t-norms can be in principle used in Zadeh's calculus but clearly their use may result in different results of the linguistic quantifier driven aggregation. It seems that the minimum operation is a good choice since it can be easily interpreted and the numerical values correspond to the intuition.

3.2 Degree of focus

The very purpose of a degree of focus is to limit the search for best linguistic summaries by taking into account some additional information in addition to the degree of truth (validity). The extended form of linguistic summaries (9) does limit by itself the search space as the search is performed in a limited subspace of all (most) trends that fulfill an additional condition specified by qualifier R. The very essence of the degree of focus introduced in this paper is to give the proportion of trends satisfying property R to all trends extracted from the time series. It provides a measure that, in addition to the basic degree of truth (validity), can help control the process of discarding nonpromising linguistic summaries. The details are described in Kacprzyk and Wilbik's paper [15].

The degree of focus is similar in spirit to a degree of covering [14], however it measures how many trends fulfill property R. That is, we focus our attention on such trends, fulfilling property R. The degree of focus makes obviously sense for the extended form summaries only, and is calculated as:

$$d_{foc}(\text{Among all } Ry_i\text{'s } Q \text{ are } P) = \frac{1}{n}\sum_{i=1}^{n}\mu_R(y_i) \qquad (12)$$

In our context, the degree of focus describes how many trends extracted from a given time series fulfill qualifier R in comparison to all extracted trends. If the degree of focus is high, then we can be sure that such a summary concerns many trends, so that it is more general. However, if the degree of focus is low, we may be sure that such a summary describes a (local) pattern seldom occurring.

As we wish to discover a more general, global relationship, we can eliminate linguistic summaries, that concern a small number of trends only. The degree of focus may be used to eliminate the whole groups of extended form summaries for which qualifier R limits the set of possible trends to, for instance, 5%. Such summaries, although they may be very true, will not be representative.

3.3 Degree of imprecision

A *degree of imprecision*, introduced by Kacprzyk and Yager in [28] and Kacprzyk, Yager and Zadrożny [29], describes how imprecise the fuzzy predicates used in the summary are. This measure does not depend on the data to be summarized, but only on the form of a summary and the definition of linguistic values.

The degree of imprecision of a single fuzzy set A_i, defining the linguistic value of a summarizer, is calculated as

$$im(A_i) = \frac{card\{x \in X_i : \mu_{A_i} > 0\}}{card X_i} \tag{13}$$

In our summaries to define membership functions of the linguistic values we use trapezoidal functions since they are sufficient in most applications [48]. Moreover, they can be very easily interpreted and defined by a user not familiar with fuzzy sets and fuzzy logic, as shown in Figure 2. To represent a fuzzy set with a trapezoidal membership function we need to store four numbers only, a, b, c and d. The use of such a form of a fuzzy set is a compromise between a so-called cointension and computational complexity (cf. Zadeh [48]).

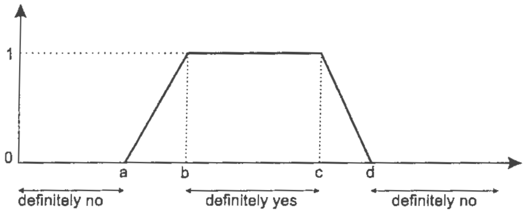

Figure 2: A trapezoidal membership function of a set

In a case of trapezoidal membership functions, defined as above, the degree of imprecision of a fuzzy set A_i is calculated as:

$$im(A_i) = \frac{d - a}{range(X_i)} \tag{14}$$

where $range(X_i)$ is the range of values taken by the feature considered.

Then, these values – calculated for each fuzzy set A_i belonging to the summarizer – are aggregated using the geometric mean. The degree of imprecision of the summary, or in fact of summarizer P, is therefore calculated as

$$im_P = \sqrt[n]{\prod_{i=1}^{n} im(A_i)} \tag{15}$$

where n is the number of fuzzy predicates in summarizer P which are defined as fuzzy sets A_i.

This degree focuses on the summarizer only. Similarly we can introduce the two additional measures, a degree of imprecision of a qualifier and that of a quantifier, as it was proposed in [36].

Hence, the degree of imprecision of a qualifier is calculated as

$$im_R = \sqrt[n]{\prod_{i=1}^{n} im(A_i)} \tag{16}$$

where n is the number of fuzzy predicates in qualifier R which are defined as fuzzy sets A_i.

And the degree of imprecision of a quantifier is calculated as

$$im_Q = im(Q) \tag{17}$$

We can aggregate those three measures using the the weighted average. Then the degree of imprecision of a simple form of the linguistic summary "Among all y's Q are P" is calculated as

$$im(\text{Among all } y\text{'s } Q \text{ are } P) = w_P im_P + w_Q im(Q) \tag{18}$$

where w_P and w_Q are the weights of the degrees of imprecision of summarizer and quantifier, respectively. $w_P, w_Q \geq 0$ and $w_P + w_Q = 1$.

The degree of imprecision of the extended form of the linguistic summary "Among all Ry'a Q are P" is calculated as

$$im(\text{Among all } Ry\text{'s } Q \text{ are } P) = w_P im_P + w_R im_R + w_Q im(Q) \tag{19}$$

where w_P, w_Q and w_R are the weights of the degrees of imprecision of summarizer, quantifier and qualifier, respectively. $w_P, w_Q, w_R \geq 0$ and $w_P + w_Q + w_R = 1$.

In the fuzzy set theory there are other concepts capturing the notion of uncertainty, like e.g. specificity or fuzziness.

3.4 Degree of specificity

The concept of specificity provides a measure of the amount of information contained in a fuzzy subset or possibility distribution. The specificity measure evaluates the degree to which a fuzzy subset points to one and only one element as its member, cf. Yager [44]. It is closely related to the inverse of the cardinality of a fuzzy set. Klir (cf. Klir and Wierman [33] or Klir and Yuan [34]) has proposed the notion of nonspecificity.

We will now consider the original Yager's proposal [44] in which the specificity measures a degree to which a fuzzy subset contains one and only one element. The measure of specificity is a measure $Sp : I^X \longrightarrow I$, $I \in [0, 1]$ if it has the following properties:

- $Sp(A) = 1$ if and only if $A = \{x\}$, (is a singleton set),
- $Sp(\varnothing) = 0$
- $\frac{\partial Sp(A)}{\partial a_1} > 0$ and $\frac{\partial Sp(A)}{\partial a_j} \leq 0$ for all $j \geq 2$

Yager [39] proposed a measure of specificity as

$$Sp(A) = \int_0^{\alpha_{max}} \frac{1}{card(A_\alpha)} d\alpha \tag{20}$$

where α_{max} is the largest membership grade in A, A_α is the α-level set of A, (i.e. $A_\alpha = \{x : A(x) \geq \alpha\}$) and $cardA_\alpha$ is the number of elements in A_α.

Let X be a continuous space, e.g. a real interval. Yager [42] proposed a general class of specificity measures in the continuous domain as

$$Sp(A) = \int_0^{\alpha_{max}} F(\mu(A_\alpha)) d\alpha \tag{21}$$

where α_{max} is the maximum membership grade in A, F is a function $F : [0,1] \longrightarrow [0,1]$ such that $F(0) = 1$, $F(1) = 0$ and $F(x) \leq F(y) \leq 0$ for $x > y$, μ is a fuzzy measure (cf. e.g. Grabisch [9]) and A_α is the α-level set.

If F is defined as $F(z) = 1 - z$, measure μ of an interval $[a, b]$ is defined as $\mu([a,b]) = b - a$, and the space is normalized to $[0,1]$, then the degree of specificity of the fuzzy set A is calculated as

$$Sp(A) = \alpha_{max} - \text{area under } A \tag{22}$$

If the fuzzy set A has a trapezoidal membership function, as e.g. shown in Figure 2, then

$$Sp(A) = 1 - \frac{c + d - (a + b)}{2} \tag{23}$$

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 (for instance, the minimum or product) for conjunction and a corresponding s-norm (for instance, the maximum or probabilistic sum, respectively) for disjunction.

We can aggregate the degrees of specificity of a summarizer, qualifier and quantifier using the weighted average. Then the degree of specificity of the simple form of the linguistic summary "Among all y's Q are P" is calculated as

$$im(\text{Among all } y\text{'s } Q \text{ are } P) = w_P Sp(P) + w_Q Sp(Q) \tag{24}$$

where w_P and w_Q are the weights of the degrees of specificity of the summarizer and quantifier, respectively. $w_P, w_Q \geq 0$ and $w_P + w_Q = 1$.

The degree of specificity of the extended form of the linguistic summary "Among all Ry's Q are P" is calculated as

$$im(\text{Among all } Ry\text{'s } Q \text{ are } P) = w_P Sp(P) + w_R Sp(R) + w_Q Sp(Q) \tag{25}$$

where w_P, w_Q and w_R are the weights of the degrees of specificity of summarizer, quantifier and qualifier, respectively. $w_P, w_Q, w_R \geq 0$ and $w_P + w_Q + w_R = 1$.

If we consider the approach proposed by Klir and his collaborators (cf. Klir and Wierman [33] or Klir and Yuan [34]) then the nonspecificity measure from fuzzy sets theory is defined using the so-called Hartley function. For a finite, nonempty (crisp) set, A, we measure this amount using a function from the class of functions

$$U(A) = c \log_b |A|, \tag{26}$$

where $|A|$ denotes the cardinality of A, b and c are positive constants, $b, c \geq 1$ (usually, $b = 2$ and $c = 1$). This function is applicable to finite sets only but it can be modified for infinite sets of $\mathbb{R}$ as follows: $U(A) = \log[1 + \mu(A)]$, where $\mu(A)$ is the measure whether 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, function $U(A)$ has the form

$$U(A) = \frac{1}{h(A)} \int_0^{h(A)} \log_2 |A^\alpha| d\alpha, \tag{27}$$

where $|A^\alpha|$ is the cardinality of the α-cut of A and $h(A)$ – the height of A. If A is a normal fuzzy set, then $h(A) = 1$.

If a nonempty fuzzy set is defined in $\mathbb{R}$ and the α-cuts are infinite sets (e.g., intervals of real numbers), then:

$$U(A) = \frac{1}{h(A)} \int_0^{h(A)} \log[1 + \mu(A^\alpha)] d\alpha, \tag{28}$$

For convenience, the values of nonspecificity are normalized.

Then the degree of specificity of "Among all y's, Q are P" may be:

$$d_s(\text{Among all } y\text{'s } Q \text{ are } P) = 1 - U(P) \tag{29}$$

and the degree of specificity of "Among all Ry's, Q are P" may be:

$$d_s(\text{"Among all } Ry\text{'s , } Q \text{ are } P\text{"}) = 1 - (U(P) \wedge U(R)) \tag{30}$$

where $U(P)$ is the degree of nonspecificity of the summarizer P, given by (28), $U(R)$ is the degree of nonspecificity of the qualifier R, and $\wedge$ is a t-norm (minimum or product).

We must emphasize the distinction between specificity and fuzziness. Fuzziness is generally related to the lack of clarity, relating to the membership of some set, whereas specificity is related to the lack og exact knowledge of some attribute.

3.5 Degree of fuzziness

A degree of fuzziness describes a degree of imprecision (which may well be equated with fuzziness) of the linguistic predicates in the summary. In general, a measure of fuzziness of a fuzzy set is a function $f : \mathcal{F} \longrightarrow \mathbb{R}^+$, where $\mathcal{F}$ denotes the family of all fuzzy subsets of X. In other words, for each fuzzy set A, this function assigns a nonnegative real number $f(A)$ that expresses a degree to which the boundary of A is not sharp.

The function f must satisfy the following three requirements (cf. Klir and Yuan [34]):

1. $f(A) = 0$ iff A is a crisp set.
2. $f(A)$ attains its maximum value iff $A(x) = 0.5$ for all $x \in X$
3. $f(A) \leq f(B)$ when set A is undoubtly sharper than set B:
   - $A(x) \leq B(x)$ when $B(x) \leq 0.5$ for all $x \in X$, or
   - $A(x) \geq B(x)$ when $B(x) \geq 0.5$ for all $x \in X$.

One way to measure the fuzziness of A is by using a distance (metric) between its membership function and the membership function of its nearest crisp set defined as: a *nearest crisp set* of a fuzzy set A is a set $\underline{A} \subset X$ given by its characteristic function:

$$\mu_{\underline{A}} = \begin{cases} 0 & \mu_A(x) \leq 0.5 \\ 1 & \mu_A(x) > 0.5 \end{cases} \tag{31}$$

Then, using different distance function we can obtain different measures, for instance:

- the linear degree of fuzziness:

$$\delta(A) = \frac{2}{n} \sum_{x \in X} |\mu_A(x_i) - \mu_{\underline{A}}(x_i)| \tag{32}$$

- the quadratic degree of fuzziness:

$$\eta(A) = \frac{2}{n}\sqrt{\sum_{x \in X}(\mu_A(x_i) - \mu_{\underline{A}}(x_i))^2} \tag{33}$$

- the vector degree of fuzziness

$$\nu(A) = \frac{2}{n}\sum_{x \in X} \mu_{\underline{A} \cap \neg A}(x_i) \tag{34}$$

Another way of measuring the (degree of) fuzziness of a fuzzy set is to measure a (degree of) lack of distinction between a fuzzy set and its complement. Of course, also here we can choose different forms of the fuzzy complements and distance functions.

If we choose the standard complement and the Hamming distance, we have:

$$f(A) = \sum x \in X(1 - |2A(x) - 1|) \tag{35}$$

where the range of f is $[0, |X|]$, $f(A) = 0$ iff A is a crisp set and $A = |X|$ when $A(x) = 0.5$ for all $x \in X$.

The above form is only valid for fuzzy sets defined in finite universes of discourse. However we can modify it to fuzzy sets defined in $\mathbb{R}$, the set of real numbers: if $X = [a, b]$, then

$$f(A) = \int_a^b (1 - |2A(x) - 1|)dx = b - a - \int_a^b |2A(x) - 1|dx \tag{36}$$

and this form of $f(.)$ will be used here.

If the set A has a trapezoidal membership function, as e.g. shown in Figure 2, then

$$f(A) = \frac{b + d - (a + c)}{2} \tag{37}$$

In general, the summarizer and the qualifier may involve more than one attribute value. To combine them we will use a t-norm (for instance, the minimum or product) for conjunction and a corresponding s-norm (for instance, the maximum or probabilistic sum, respectively) for the disjunction.

The degree of fuzziness of "Among all y's, Q are P" is:

$$d_f(\text{Among all } y\text{'s } Q \text{ are } P) = f(P) \wedge f(Q) \tag{38}$$

where $f(P)$ is the degree of fuzziness of the summarizer P, $f(Q)$ is the degree of fuzziness of the quantifier Q, and $\wedge$ is a t-norm (minimum or product).

The degree of fuzziness of "Among all Ry's, Q are P" is:

$$d_f(\text{Among all } Ry\text{'s } Q \text{ are } P) = f(P) \wedge f(R) \wedge f(Q) \tag{39}$$

where $f(P)$ is the degree of fuzziness of the summarizer P, $f(R)$ is the degree of fuzziness of the qualifier R, $f(Q)$ is the degree of fuzziness of the quantifier Q, and $\wedge$ is a t-norm (minimum or product).

The degree of fuzziness is not of high importance in evaluation of the summaries. However we discussed it for completeness.

4 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 January 2002 until the end March 2009 with the value of one share equal to PLN 12.06 in the beginning of the period to PLN 21.82 at the end of the time span considered (PLN stands for the Polish Zloty). The minimal value recorded was PLN 9.35 while the maximal one during this period was PLN 57.85. The biggest daily increase was equal to PLN 2.32, while the biggest daily decrease was equal to PLN 3.46.

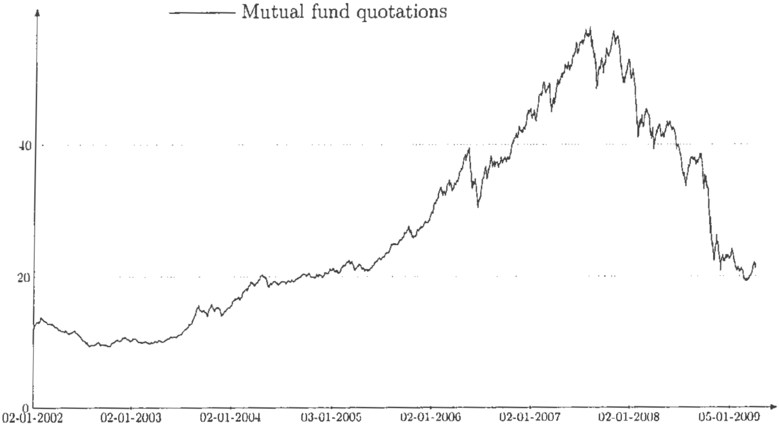

Figure 3: Mutual fund quotations

It should be noted that the example shown below is meant to illustrate the method proposed by analyzing the absolute performance of a given investment fund.

We do not deal here with a presumably more common way of analyzing an investment fund by relating its performance to a benchmark (or benchmarks) exemplified by an average performance of a group of (similar) funds, a stock market index or a synthetic index reflecting, for instance, the bond versus stock allocation.

Using the modified Sklansky and Gonzalez algorithm (cf. [38]) and $\varepsilon = 0.25$ we obtained 422 extracted trends. The shortest trend took 1 time unit only, while the longest one – 71. The histograms for duration, dynamics of change and variability are shown in Figure 4.

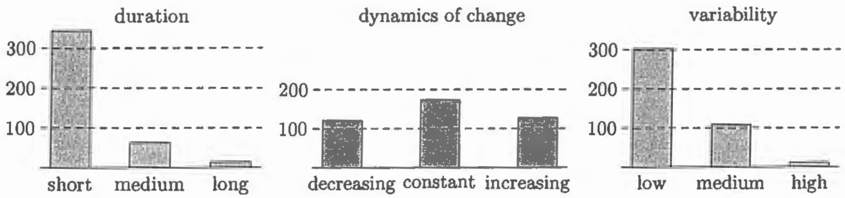

Figure 4: Histograms of duration, dynamics of change and variability

We have applied different granulations, namely with 3, 5 and 7 labels for each feature (dynamics of change, duration and variability). Minimal accepted truth value was 0.6 and the degree of focus threshold was 0.1. The degree of focus, and the method of effective and efficient generating summaries is described in Kacprzyk and Wilbik's paper [15].

If we have used 3 labels for dynamics of change (decreasing, constant and increasing), 3 labels for duration (short, medium length and long) and 3 labels for variability (low, moderate and high) we have obtained the summaries shown in Table 1.

The linguistic summaries are sorted according to the truth values, and later by the values of degree of focus. The simple form summaries are before the extended ones with the same truth value. The summaries here have high values of specificity, indicating that they may be potentially useful for the user. Only a few summaries have the degree of imprecision greater than 0.5, and they should be analyzed with care. The values of degree of focus are small, only for 3 summaries they exceed the value of 0.2.

Let us now slightly modify the used properties. We add linguistic labels A, B, C. Their membership functions together with the membership function of the fuzzy set with label *low* are depicted in Fig. 5.

Table 1: Results for 3 labels

linguistic summary	truth value	degree of focus	degree of imprecision	degree of specificity	degree of fuzziness
Among all y's, most are short	1		0.385	0.745	0.135
Among all low y's, most are short	1	0.7227	0.39	0.73	0.1567
Among all increasing y's, most are short	1	0.2984	0.4047	0.6867	0.0993
Among all increasing y's, almost all are short	1	0.2984	0.2713	0.77	0.0493
Among all decreasing y's, most are short	1	0.2880	0.4047	0.6867	0.0993
Among all decreasing y's, most are short and low	1	0.2880	0.4371	0.6067	0.0993
Among all decreasing y's, most are low	1	0.2880	0.5147	0.6067	0.1593
Among all decreasing y's, almost all are short	1	0.2880	0.2713	0.77	0.0493
Among all short and decreasing y's, most are low	1	0.2842	0.4254	0.6067	0.1567
Among all medium y's, most are constant	1	0.1308	0.3393	0.765	0.1253
Among all low y's, almost all are short	0.9674	0.7227	0.2567	0.8133	0.1067
Among all increasing y's, most are low	0.9610	0.2984	0.5147	0.6067	0.1593
Among all short and increasing y's, most are low	0.9588	0.2946	0.4254	0.6067	0.1567
Among all short y's, most are low	0.9483	0.8341	0.39	0.73	0.1567
Among all increasing y's, most are short and low	0.9386	0.2984	0.4371	0.6067	0.0993
Among all y's, most are low	0.8455		0.55	0.625	0.225
Among all decreasing y's, almost all are low	0.8122	0.2880	0.3813	0.69	0.1093
Among all decreasing y's, almost all are short and low	0.7916	0.2880	0.3038	0.69	0.0493
Among all moderate y's, most are short	0.7393	0.2483	0.4567	0.6967	0.2233
Among all short and constant y's, most are low	0.7325	0.2565	0.4028	0.7033	0.1567
Among all moderate y's, most are constant	0.7024	0.2483	0.4893	0.67	0.2353
Among all y's, most are short and low	0.6915		0.4337	0.625	0.135
Among all y's, almost all are short	0.6706		0.185	0.87	0.06
Among all constant y's, most are short	0.6405	0.4136	0.3127	0.7833	0.1087

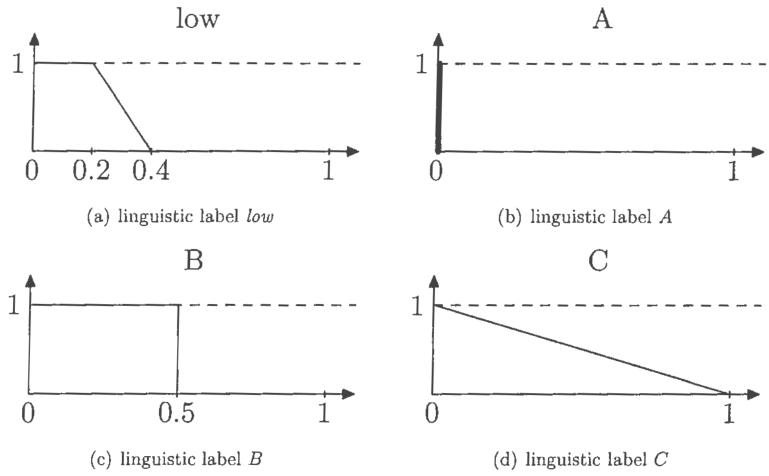

Figure 5: Illustration of the membership functions for the linguistic labels *low*, *A*, *B* and *C*.

The values of imprecision, specificity and fuzziness of a single fuzzy set are shown in Table 2.

Table 2: The values of imprecision, specificity and fuzziness of fuzzy sets representing linguistic labels *low*, *A*, *B* and *C*

lingistic label	imprecision	specificity	fuzziness
low	0.4	0.7	0.2
A	0.006	0.994	0
B	0.5	0.5	0
C	1	0.497	0.994

Let us now analyze some of the summaries:

Table 3: Some of the obtained summaries with the linguistic labels *low*, *A*, *B* and *C*

linguistic summary	truth value	imprecision	specificity	fuzziness
Among all y's, most are *low*	0.8455	0.55	0.625	0.225
Among all y's, most are *A*	0.5280	0.353	0.772	0.125
Among all y's, most are *B*	1	0.6	0.525	0.125
Among all y's, most are *C*	1	0.85	0.5235	0.622

Here, (Tab. (3)), we may observe, that higher values of the degree of specificity (or lower of the degree of imprecision) may result in lower truth values. The low or high values of fuzziness do not considerably affect the results in this case.

Those big differences in the values of degrees of imprecision, specificity and fuzziness are visible also in evaluation of more complex linguistic summaries, e.g. Tab. (4). The truth values and the values of degrees of focus are not shown in Table (4) as they are the same for all 4 summaries and equal 1.0 and 0.2842, respectively.

Table 4: Some of the obtained summaries with the linguistic labels *low*, A, B and C

linguistic summary	imprecision	specificity	fuzziness
Among all short and decreasing y's, most are *low*	0.4254	0.6067	0.1567
Among all short and decreasing y's, most are A	0.2941	0.7047	0.09
Among all short and decreasing y's, most are B	0.4588	0.54	0.09
Among all short and decreasing y's, most are C	0.6254	0.539	0.4213

The high values of imprecision of the last two summaries indicate, that they are general, and should be analyzed with special care. On the other hand, high values of the degree of specificity indicate, that those summaries may be promising and useful.

Similar observations may me made if the modified label is used as the qualifier, e.g. Table (5). We can easily see the change in the values of the degree of focus, so the different number of data described by our summaries. Here again the truth value is equal 1.0 for all 4 summaries.

Table 5: Some of the obtained summaries with the linguistic labels *low*, A, B and C

linguistic summary	focus	imprecision	specificity	fuzziness
Among all *low* y's, most are short	0.7227	0.39	0.73	0.1567
Among all A y's, most are short	0.5640	0.2587	0.828	0.09
Among all B y's, most are short	0.8910	0.4233	0.6633	0.09
Among all C y's, most are short	0.8353	0.59	0.6623	0.4213

Again relatively high values of imprecision of the last two summaries indicate, that they are general, and should be analyzed with special care. Very high values of the degree of specificity indicate, that those summaries may be promising and useful.

If we have used 5 labels for dynamics of change (quickly decreasing, decreasing, constant, increasing and quickly increasing), 5 labels for duration (very short, short, medium length, long and very long) and 5 labels for variability (very low, low, moderate, high and very high) we have obtained the summaries shown in Table 6.

Similarly, the linguistic summaries are sorted according to the truth values, and later by the values of degree of focus. The summaries here have high values of degree of specificity, indicating that they may be potentially useful for the user. Values of the degree of specificity are higher than the values than in the case with 3 linguistic labels. Values of the degree of imprecision as well as the ones of the degree of focus are smaller than in the case with 3 linguistic labels.

If we have used 7 labels for dynamics of change (quickly decreasing, decreasing, slowly decreasing, constant, slowly increasing, increasing and quickly increasing), 7 labels for duration (very short, short, rather short, medium length, rather long, long and very long) and 7 labels for variability (very low, low, rather low, moderate, rather high, high and very high) we have obtained the summaries shown in Table 7.

Similarly, the linguistic summaries are sorted according to the truth values, and later by the values of degree of focus. There are only 4 summaries that describe "global" situation. Those are one simple form and 3 with high values of the degree of focus. The degree of focus for the other summaries is smaller than 15%, so they describe patterns more locally occurring. The summaries here have high values of degree of specificity, indicating that they may be potentially useful for the user. Values of the degree of imprecision as well as the ones of the degree of focus are small.

Let us note that the degree of imprecision and the degree of specificity are to some extent related, notably large values of the degree of imprecision are associated with small values of the degree of specificity and vice versa. The degree of focus describes a different aspect. So it is possible to have a summary with a very small value of the degree of specificity (i.e. big of the degree of imprecision) which may have either a very small or big value of the degree of fuzziness. However, if a summary has a very high degree of specificity, then its degree of focus is low.

5 Concluding remarks

We extended our approach to the linguistic summarization of time series based on a calculus of linguistically quantified propositions used for a linguistic quantifier

Table 6: Results for 5 labels

linguistic summary	truth value	degree of focus	degree of imprecision	degree of specificity	degree of fuzziness
Among all very short y's, most are very low	1	0.7180	0.3167	0.7867	0.1233
Among all very low y's, most are very short	1	0.6141	0.3167	0.7867	0.1233
Among all very low y's, almost all are very short	1	0.6141	0.1833	0.87	0.0733
Among all increasing y's, most are very short	1	0.1903	0.2967	0.7993	0.1087
Among all quickly decreasing y's, most are very short	1	0.1484	0.3613	0.73	0.0933
Among all quickly decreasing y's, most are very short and very low	1	0.1484	0.378	0.6933	0.0933
Among all quickly decreasing y's, most are very low	1	0.1484	0.4113	0.6933	0.126
Among all quickly decreasing y's, almost all are very short	1	0.1484	0.228	0.8133	0.04933
Among all quickly decreasing y's, almost all are very short and very low	1	0.1484	0.2447	0.7767	0.0493
Among all quickly decreasing y's, almost all are very low	1	0.1484	0.278	0.7767	0.076
Among all very short and quickly decreasing y's, most are very low	1	0.1464	0.3431	0.6933	0.1233
Among all decreasing y's, most are very short	1	0.1434	0.2967	0.7993	0.1087
Among all very short and decreasing y's, most are very low	1	0.1275	0.3279	0.7627	0.1233
Among all quickly increasing y's, most are very short	1	0.1101	0.3613	0.73	0.0933
Among all quickly increasing y's, most are very short and very low	1	0.1101	0.378	0.6933	0.0933
Among all quickly increasing y's, most are very low	1	0.1101	0.4113	0.6933	0.126
Among all quickly increasing y's, almost all are very short	1	0.1101	0.228	0.8133	0.0493
Among all quickly increasing y's, almost all are very short and very low	1	0.1101	0.2447	0.7767	0.04933
Among all quickly increasing y's, almost all are very low	1	0.1101	0.278	0.7767	0.076
Among all very short and quickly increasing y's, most are very low	1	0.1100	0.3431	0.6933	0.1233
Among all decreasing y's, almost all are very short	0.9446	0.1434	0.1633	0.8827	0.0587
Among all short y's, most are constant	0.8999	0.1979	0.3193	0.78	0.1153
Among all low y's, most are constant	0.8872	0.1471	0.3893	0.7367	0.1687
Among all decreasing y's, most are very low	0.8585	0.1434	0.3467	0.7627	0.1353
Among all decreasing y's, most are very short and very low	0.8477	0.1434	0.3133	0.7627	0.1087
Among all y's, most are very short	0.8360		0.375	0.755	0.135
Among all very short and increasing y's, most are very low	0.7720	0.1611	0.3279	0.7627	0.1233
Among all very short and constant y's, most are very low	0.7572	0.1857	0.3306	0.7533	0.1233
Among all increasing y's, almost all are very short	0.7325	0.1903	0.1633	0.8827	0.0587
Among all moderate y's, most are constant	0.6674	0.1987	0.4227	0.7033	0.1687
Among all y's, most are very low	0.6282		0.45	0.7	0.175

Table 7: Results for 7 labels

linguistic summary	truth value	degree of focus	degree of imprecision	degree of specificity	degree of fuzziness
Among all very low y's, most are very short	1	0.5984	0.3	0.795	0.1067
Among all very low y's, almost all are very short	1	0.5984	0.1667	0.8783	0.0567
Among all quickly decreasing y's, most are very short	1	0.1484	0.3613	0.73	0.0933
Among all quickly decreasing y's, most are very short and very low	1	0.1484	0.3735	0.7017	0.0933
Among all quickly decreasing y's, most are very low	1	0.1484	0.3947	0.7017	0.1093
Among all quickly decreasing y's, almost all are very short	1	0.1484	0.228	0.8133	0.0493
Among all quickly decreasing y's, almost all are very short and very low	1	0.1484	0.2402	0.785	0.0493
Among all quickly decreasing y's, almost all are very low	1	0.1484	0.2613	0.785	0.0593
Among all very short and quickly decreasing y's, most are very low	1	0.1464	0.3264	0.7017	0.1067
Among all increasing y's, most are very short	1	0.1345	0.2873	0.8087	0.1087
Among all slowly decreasing y's, most are very short	1	0.1124	0.278	0.818	0.1087
Among all quickly increasing y's, most are very short	1	0.1101	0.3613	0.73	0.0933
Among all quickly increasing y's, most are very short and very low	1	0.1101	0.3735	0.7017	0.0933
Among all quickly increasing y's, most are very low	1	0.1101	0.3947	0.7017	0.1093
Among all quickly increasing y's, almost all are very short	1	0.1101	0.228	0.8133	0.0493
Among all quickly increasing y's, almost all are very short and very low	1	0.1101	0.2402	0.785	0.0493
Among all quickly increasing y's, almost all are very low	1	0.1101	0.2613	0.785	0.0593
Among all very short and quickly increasing y's, most are very low	1	0.1100	0.3264	0.7017	0.1067
Among all very short y's, most are very low	0.9847	0.7180	0.3	0.795	0.1067
Among all y's, most are very short	0.8360		0.375	0.755	0.135
Among all very short and increasing y's, most are very low	0.7938	0.11533	0.3083	0.7803	0.1067
Among all increasing y's, almost all are very short	0.7881	0.1345	0.154	0.892	0.0587
Among all slowly increasing y's, most are very short	0.7591	0.1372	0.278	0.818	0.1087
Among all slowly decreasing y's, most are very low	0.6361	0.1124	0.3113	0.7897	0.1187

driven aggregation of partial scores (trends). We presented a reformulation and extension of our works mainly by including a more complex evaluation of the linguistic summaries obtained. In addition to the degree of truth (validity), we additionally used a degree of imprecision, specificity, fuzziness and focus. However, for simplicity and tractability, we used in the first shot the degrees of truth (validity) and focus, to reduce the space of possible linguistic summaries, and then – for a usually much smaller set of linguistic summaries obtained – we used the remaining four degrees of imprecision, specificity and fuzziness for making a final choice of appropriate linguistic summaries. So this does not guarantee the optimality, our experience however suggests that it makes possible to generate good summaries in computationally reasonable time. A more formalized approach of this heuristic method to find best summaries will be presented in next papers. We showed an application to the absolute performance type analysis of daily quotations of an investment fund.

References

[1] A 10-step guide to evaluating mutual funds. www.personalfn.com/detail.asp?date=5/18/2007&story=2.

[2] New year's eve:past performance is no indication of future return. stockcasting.blogspot.com/2005/12/new-years-evepast-performance-is-no.html.

[3] Past performance does not predict future performance. www.freemoneyfinance.com/2007/01/past_performanc.html.

[4] Past performance is not everything. www.personalfn.com/detail.asp?date=9/1/2007&story=3.

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

[6] Ildar Batyrshin and Leonid Sheremetov. Perception based functions in qualitative forecasting. In Ildar Batyrshin, Janusz Kacprzyk, Leonid Sheremetov, and Lotfi A. Zadeh, editors, *Perception-based Data Mining and Decision Making in Economics and Finance*. Springer-Verlag, Berlin and Heidelberg, 2006.

[7] John C. Bogle. *Common Sense on Mutual Funds: New Imperatives for the Intelligent Investor*. Wiley, New York, 1999.

[8] Patric Bosc, Ludovic Lietard, and Olivier Pivet. Quantified statements and database fuzzy queries. In Patric Bosc and Janusz Kacprzyk, editors, *Fuzziness in Database Management Systems*. Springer-Verlag, Berlin and Heidelberg, 1995.

[9] Michel Grabisch. Fuzzy integral as a flexible and interpretable tool of aggregation. In Bernadette Bouchon-Meunier, editor, *Aggregation and Fusion of Imperfect Information*, pages 51–72. Heidelberg, New York: Physica–Verlag, 1998.

[10] Janusz Kacprzyk and Sławomir Zadrożny. Fuzzy linguistic data summaries as a human consistent, user adaptable solution to data mining. In Bogdan Gabrys, Kauko Leiviska, and Jens Strackeljan, editors, *Do Smart Adaptive Systems Exist?*, pages 321–339. Springer, Berlin, Heidelberg, New York, 2005.

[11] Janusz Kacprzyk and Sławomir Zadrożny. Linguistic database summaries and their protoforms: toward natural language based knowledge discovery tools. *Information Sciences*, 173:281–304, 2005.

[12] Janusz Kacprzyk and Anna Wilbik. An extended, specificity based approach to linguistic summarization of time series. In *Proceedings of the 12th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 551–559, 2008.

[13] Janusz Kacprzyk and Anna Wilbik. Linguistic summarization of time series using linguistic quantifiers: augmenting the analysis by a degree of fuzziness. In *Proceedings of 2008 IEEE World Congress on Computational Intelligence*, pages 1146–1153. IEEE Press, 2008.

[14] Janusz Kacprzyk and Anna Wilbik. A new insight into the linguistic summarization of time series via a degree of support: Elimination of infrequent patterns. In D. Dubois, M.A. Lubiano, H. Prade, Maria A. Gil, Przemysaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Soft Methods for Handling Variability and Imprecision*, pages 393–400. Springer-Verlag, Berlin and Heidelberg, 2008.

[15] Janusz Kacprzyk and Anna Wilbik. Towards an efficient generation of linguistic summaries of time series using a degree of focus. In *Proceedings of the 28th North American Fuzzy Information Processing Society Annual Conference – NAFIPS 2009*, 2009.

[16] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Capturing the essence of a dynamic behavior of sequences of numerical data using elements of a quasi-natural language. In *Proceedings of the "2006 IEEE International Conference on Systems, Man, and Cybernetics"*, pages 3365–3370. IEEE Press, 2006.

[17] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. A linguistic quantifier based aggregation for a human consistent summarization of time series. In Jonathan Lawry, Enrique Miranda, Alberto Bugarin, Shoumei Li, Maria A. Gil, Przemysaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Soft Methods for Integrated Uncertainty Modelling*, pages 186–190. Springer-Verlag, Berlin and Heidelberg, 2006.

[18] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Linguistic summaries of time series via a quantifier based aggregation using the Sugeno integral. In *Proceedings of 2006 IEEE World Congress on Computational Intelligence*, pages 3610–3616. IEEE Press, 2006.

[19] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Linguistic summarization of trends: a fuzzy logic based approach. In *Proceedings of the 11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, pages 2166–2172, 2006.

[20] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. On some types of linguistic summaries of time series. In *Proceedings of the 3rd International IEEE Conference "Intelligent Systems"*, pages 373–378. IEEE Press, 2006.

[21] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Analysis of time series via their linguistic summarization: the use of the sugeno integral. In *Proceedings of the 7th International Conference on Intelligent Systems Design and Applications - ISDA 2007*, pages 262–267. IEEE Press, 2007.

[22] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Linguistic summaries of time series via an owa operator based aggregation of partial trends. In *Proceedings of the FUZZ-IEEE 2007 IEEE International Conference on Fuzzy Systems*, pages 467–472. IEEE Press, 2007.

[23] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Linguistic summarization of time series by using the choquet integral. In Patricia Melin, Oscar Castillo, Luis T. Aguilar, Janusz Kacprzyk, and Witold Pedrycz, editors, *Foundations of Fuzzy Logic and Soft Computing - IFSA 2007*, pages 284–294. Springer-Verlag, Berlin and Heidelberg, 2007.

[24] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Linguistic summarization of time series under different granulation of describing features. In Marzena Kryszkiewicz, James F. Peters, Henryk Rybinski, and Andrzej Skowron, editors, *Rough Sets and Intelligent Systems Paradigms - RSEISP 2007*, pages 230–240. Springer-Verlag, Berlin and Heidelberg, 2007.

[25] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Mining time series data via linguistic summaries of trends by using a modified sugeno integral based aggregation. In *IEEE Symposium on Computational Intelligence and Data Mining CIDM 2007*, pages 467–472. IEEE Press, 2007.

[26] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. On linguistic summaries of time series via a quantifier based aggregation using the sugeno integral. In Oscar Castillo Patricia Melin, Janusz Kacprzyk, and Witold Pedrycz, editors, *Hybrid Intelligent Systems Analysis and Design*, pages 421–439. Springer-Verlag, Berlin and Heidelberg, 2007.

[27] Janusz Kacprzyk, Anna Wilbik, and Sławomir Zadrożny. Linguistic summarization of time series using a fuzzy quantifier driven aggregation. *Fuzzy Sets and Systems*, 159(12):1485–1499, 2008.

[28] Janusz Kacprzyk and Ronald R. Yager. Linguistic summaries of data using fuzzy logic. *International Journal of General Systems*, 30:33–154, 2001.

[29] Janusz Kacprzyk, Ronald R. Yager, and Sławomir Zadrożny. A fuzzy logic based approach to linguistic summaries of databases. *International Journal of Applied Mathematics and Computer Science*, 10:813–834, 2000.

[30] Janusz Kacprzyk, Ronald R. Yager, and Sławomir Zadrożny. Fuzzy linguistic summaries of databases for an efficient business data analysis and decision support. In J. Zurada W. Abramowicz, editor, *Knowledge Discovery for Business Information Systems*, pages 129–152. Kluwer, Boston, 2001.

[31] Eammon Keogh, Selina Chu, David Hart, and Michael Pazzani. An online algorithm for segmenting time series. In *Proceedings of the 2001 IEEE International Conference on Data Mining*, 2001.

[32] Eammon Keogh, Selina Chu, David Hart, and Michael Pazzani. Segmenting time series: A survey and novel approach. In M. Last, A. Kandel, and H. Bunke, editors, *Data Mining in Time Series Databases*. World Scientific Publishing, 2004.

[33] George J. Klir and Mark J. Wierman, editors. *Uncertainty-Based Information, Elements of Generalized Information Theory.* Physica-Verlag, 1999.

[34] George J. Klir and Bo Yuan, editors. *Fuzzy Stes and Fuzzy Logic, Theory and Applications.* Prentice Hall, 1995.

[35] Randy Myers. Using past performance to pick mutual funds. *Nation's Business*, Oct, 1997. findarticles.com/p/articles/ mi_m1154/is_n10_v85/ai_19856416.

[36] Adam Niewiadomski, editor. *Methods for the Linguistic Summarization of Data: Aplications of Fuzzy Sets and Their Extensions.* Academic Publishing House EXIT, 2008.

[37] U.S. Securities and Exchange Commission. Mutual fund investing: Look at more than a fund's past performance. www.sec.gov/investor/pubs/mfperform.htm.

[38] Jack Sklansky and Victor Gonzalez. Fast polygonal approximation of digitized curves. *Pattern Recognition*, 12(5):327–331, 1980.

[39] Ronald R. Yager. Measuring tranquility and anxiety in decision making: An application of fuzzy sets. *International Journal of General Systems*, 8:139–146, 1982.

[40] Ronald R. Yager. A new approach to the summarization of data. *Information Sciences*, 28:69–86, 1982.

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

[42] Ronald R. Yager. On the specificity of a possibility distribution. *Fuzzy Sets and Systems*, 50:279–292, 1992.

[43] Ronald R. Yager. Quantifier guided aggregation using OWA operators. *International Journal of Intelligent Systems*, 11:49–73, 1996.

[44] Ronald R. Yager. On measures of specificity. In Okyay Kaynak, Lotfi A. Zadeh, Burhan Trksen, and Imre J. Rudas, editors, *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications*, pages 94–113. Springer-Verlag: Berlin, 1998.

[45] Ronald R. Yager and Janusz Kacprzyk, editors. *The Ordered Weighted Averaging Operators: Theory and Applications.* Kluwer, Boston, 1997.

[46] Lotfi A. Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, 9(2):111–127, 1983.

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

[48] Lotfi A. Zadeh. Computation with imprecise probabilities. In *In IPMU'08, Torremolinos (Malaga), June 22-27, 2008*, 2008.