INFORMATICA, 2022, Vol. 33, No. 2, 321–342

DOI: https://doi.org/10.15388/22-INFOR492

A New Quality Measure and Visualization of the Short-Quantified Sentences of Natural Language on Maps – A Case on COVID-19 Data

Miroslav HUDEC[1,3,*], Kristína MALOVCOVÁ[2], Rijad TRUMIC[1], Eva RAKOVSKÁ[3]

[1] *Faculty of Economics, VSB – Technical Unversity of Ostrava, Sokolská třída 33, Ostrava, Czech Republic*
[2] *DÔVERA zdravotná poisťovňa, a.s., Bratislava, Slovak Republic*
[3] *Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, Bratislava, Slovak Republic*
e-mail: miroslav.hudec@vsb.cz, malovcova.kristina@dovera.sk, rijad.trumic@vsb.cz, eva.rakovska@euba.sk

Received: July 2021; accepted: June 2022

Abstract. Maps are a common tool for visualizing various statistical figures that describe development in our society. Domain experts, journalists, and general public can pose questions on how to emphasize regions where, for instance, *most young patients have long stayed in hospitals*. One of the visualization's problems is expressing validities of short-quantified sentences for regions on maps. The truth value of a summary assigns a value from the unit interval, which makes it suitable for interpretation on maps by hues of a selected colour, but it does not reflect the data distribution among regions. To meet this goal, a new quality measure covering data distribution among districts and its aggregation by the ordinal sums of conjunctive and disjunctive functions with the truth value is proposed and documented on examples. The next proposal is a relative quantifier expressing *significant proportion* of entities. This model is applied to the interpretation of COVID-19 cases development in the Slovak Republic on real data from one health insurance company. Finally, this article discusses the applicability of the proposed approach in other areas where the interpretation of summarized sentences on maps is beneficial.

Key words: linguistic summary, visualization on maps, quality measure of summary, aggregation function, relative quantifier, COVID-19.

1. Introduction

Public administrations, domain experts, journalists, and the public are generally interested in the data and information that describe various aspects of our society. Currently, they are interested more in the data and information regarding the COVID-19 or environmental problems. Such information is often related to territorial units and therefore

*Corresponding author.

could be explained on maps. The visualization of statistical figures and graphs on maps is a well-established field. Users ranked from the domain experts to the general audience get an overview of the distribution of the considered figures among territorial units at first glance. However, statistical figures are comprehensible only for users having a considerable level of statistical literacy (Hudec *et al.*, 2018). Charts are also a common way to visually interpret data and relationships, but require training and experience to be interpreted quickly and accurately (Reiter, 2017). An intelligent visualization on maps should understand user's information-seeking goals, i.e. possessing the capability to select, process, and visualize the relevant data in a way that is productive for achieving user's goals (Paliulionis, 2000).

Maps are a common sight in official statistics data dissemination and visualization (e.g. GDP, poverty, income, and currently COVID-19). Figures are usually aggregated from the lower territorial levels (for instance, the amount of produced waste in districts is calculated from their respective municipalities) or institutions (e.g. the number of people with a particular diagnosis from all medical doctors' reports in a region). In the European Union, countries are divided into three NUTS (Nomenclature des Unités Territoriales Statistiques) levels and two LAU (Local Administrative Unit) levels (Eurostat, 2003). The first three levels have a unique code assigned by Eurostat. It makes the comparison of these units among the EU countries easier.

A single statistical figure can be interpreted by hues of a selected colour reflecting the values for each region. Pie charts are applied when the proportion of several values should be visualized (e.g. votes in elections by parties for each region). However, the problem is interpreting dependencies among several attributes when sharp boundaries of the considered categories cannot be constructed, or a natural uncertainty prevails. For instance, to visualize regions, where *the most of low-altitude municipalities have high pollution*. The terms *most of*, *low-altitude*, and *high pollution* are intuitively clear. The same holds for evaluating regions with a sentence like *the most of young customers buy groceries in the early evening*. However, we are not able to clearly express which moment separates the early evening from the late evening. The same holds for other adjectives and linguistic quantifiers. For the quantifier *most of*, a higher proportion of entities satisfying a condition means a higher truth value of the sentence.

This structure of Linguistic Summaries (LSs) has been proposed by Yager (1982). Since then, summaries have been significantly advanced (Kacprzyk and Yager, 2001; Kacprzyk and Zadrożny, 2005; Kacprzyk *et al.*, 2006; Liétard, 2006; Wilbik *et al.*, 2020). The truth value of an LS can be false (value 0), true (value 1), or true to some extent (a value from the open unit interval) and therefore it is interpretable on maps. To avoid summaries based on the outliers or low data coverage, quality measures have been developed in, e.g. (Hudec, 2017; Kacprzyk and Strykowski, 1999; Wu *et al.*, 2010). Quality measures usually focus on a single summary or a set of summaries to recognize the most representative ones (Bugarín *et al.*, 2015).

The use of linguistic summaries provides a way for verbalizing data mining tasks by a graphical interface contributing to the interpretation of results (Kacprzyk and Zadrożny, 2010). The research questions of this work are the following: Could we interpret the validity of a single LS for all districts on maps covering the data distributions among districts, and for this purpose develop a new quality measure to calculate the validity of the same summarized sentence for each district? By this approach, we can bridge the research gap in evaluating the quality of a summary for diverse regions and interpret the intensities of summaries' validities on maps.

The first novelty is a new quality measure covering data distribution among districts and integrating it with the truth value by an aggregation function of the mixed behaviour. The second novelty is applying it on real-world data to provide a practical solution for explaining COVID-19 cases in all districts of the Slovak Republic on maps. The compactness and robustness of the proposed solution make it interesting for other medical data as well as for environmental or business data.

The remainder of this article is organized as follows: Section 2 introduces linguistic summaries and related concepts. Section 3 is dedicated to the new quality measure for summaries on the subsets of data. Section 4 explains the interpretation of summaries on maps by the proposed method, whereas Section 5 is devoted to the experiments on real-data. Section 6 discusses the solution and its applicability. Finally, Section 7 answers the research questions and concludes the article.

2. Preliminaries of Linguistic Summaries and Quality Measures

The field of linguistic summarization splits into several directions: classic prototype forms, temporal summaries, summaries of time series, summaries of textual data. A brief overview can be found in Boran *et al.* (2016), Lesot *et al.* (2016). In this work, we focus on the classic prototype forms initially proposed by Yager (1982) as *Q entities in* $\mathbf{X}$ *are P* (e.g. *a significant proportion of patients is young*) and *Q R entities in* $\mathbf{X}$ *are P* (e.g. *a significant proportion of young patients have long stayed in hospitals*). In this context, Q is a relative quantifier, P and R are adjectives also formalized by fuzzy sets, whereas $\mathbf{X}$ is a collection of entities, in which each entity is explained by a set of attributes. More information about classic prototype forms and their applications is in Hudec *et al.* (2018), Kacprzyk and Zadrożny (2005), Kacprzyk *et al.* (2006), Liétard (2006), Rasmussen and Yager (1997), Wilbik *et al.* (2020), Wu *et al.* (2010).

2.1. *Classic Prototype Forms*

Linguistic summaries rely on the theory of fuzzy sets and fuzzy logic, where belonging to a set is a matter of degree. A fuzzy set F is defined by a membership function μ_F, a mapping from the universe of discourse X to a completely distributive lattice (in this case unit interval) that matches each element of X with its degree of membership to the set F (Zadeh, 1965)

$$\mu_F(x) : X \to [0, 1], \tag{1}$$

where $\mu_F(x) = 0$ means that an element x clearly does not belong to F, while $\mu_F(x) = 1$ means that x is a full member of F. A value of $\mu_F(x)$ between 0 and 1 indicates the intensity by which the element $x \in X$ belongs to F.

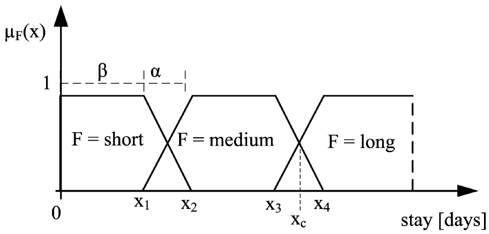

Fig. 1. Linguistic variable *length of stay* and its labels.

The next concept required for this work is Linguistic Variable (LV). It is a variable, whose values (often called labels) are words of natural language determined by a quintuple $(L, G(L), X, M, H)$ (Zadeh, 1975), where:

- L is the name of the variable,
- $G(L)$ is a set of all linguistic labels related to variable L,
- X is the universe of discourse,
- M is the syntactic rule for generating $G(L)$ values,
- H is the semantic rule that relates each linguistic label of $G(L)$ to its meaning $H(L)$.

A LV should also be a fuzzy partition (Ruspini, 1969; Alonso *et al.*, 2021) to ensure that the sum of matching degrees to all sets is equal to 1 (usually to the neighbouring ones). An example of LV is any attribute whose domain can be divided into overlapping granules, e.g. *length of stay*. The LV *length of stay* consisting of labels *short*, *medium*, and *long* is plotted in Fig. 1, where α is the uncertainty area between the two neighbouring sets and β is the length of the fuzzy set core, i.e. values which fully belong to the set. Observe that, when $\alpha = 0$, we get the classical intervals, whereas for $\beta = 0$, we get the maximal uncertainty in concepts.

The syntactic rule explains the required number of linguistic labels and their names (in this case three, but a finer or coarser granularity can be created), whereas the semantic rule assigns the context-dependent meaning to each label by fuzzy sets. Generally, the fuzzy set *long* (or other terms expressing larger amounts like *high* or *big*) is expressed as an increasing function. In this work, we adopted the linear functions due to their simplicity, but non-linear ones can be straightforwardly applied (Holzinger *et al.*, 2017; Hudec *et al.*, 2018). In this context, the fuzzy set *long* is formalized as (see Fig. 1):

$$\mu_{long}(x) = \begin{cases} 1 & \text{for } x \geqslant x_4, \\ \frac{x - x_4}{x_4 - x_3} & \text{for } x_3 < x < x_4, \\ 0 & \text{for } x \leqslant x_3. \end{cases} \tag{2}$$

Value x_c in Fig. 1 is the maximal uncertainty point. In a smooth transition from sets *medium* to *long*, x_c belongs to both with 0.5 degree.

The next key element in LSs is the fuzzy relative quantifier. In this work, we adopted the sigma-counts approach (Zadeh, 1983) for its simplicity. In this way, all building blocks of LSs are modelled by the same approach, which makes the whole process effective, especially for visualization on maps. Within that approach, the proportional non-decreasing quantifier is formalized by a function where $\mu_Q(0) = 0$ and $\mu_Q(1) = 1$ (Kacprzyk and Zadrożny, 2005) as

$$\mu_Q(y) = \begin{cases} 1 & \text{for } y \geqslant 0.8, \\ \frac{y-0.3}{0.5} & \text{for } 0.3 < y < 0.8, \\ 0 & \text{for } y \leqslant 0.3. \end{cases} \tag{3}$$

When formalizing quantifier *most of*, the non-decreasing function starts to increase in 0.5, to cover the natural meaning of majority, i.e.

$$\mu_Q(y) = \begin{cases} 1 & \text{for } y \geqslant 0.8, \\ \frac{y-0.5}{0.3} & \text{for } 0.5 < y < 0.8, \\ 0 & \text{for } y \leqslant 0.5. \end{cases} \tag{4}$$

By the above defined fuzzy sets and the quantifier, we can calculate the truth value of LS of the form Q *entities in* $\mathbf{X}$ *are* P (we call it the basic structure) as Yager (1982)

$$v_{LSb}(\mathbf{X}) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(x_i)\right), \tag{5}$$

where n is the number of entities in a considered data set $\mathbf{X}$, and membership function μ formalizes quantifier Q and predicate P. The truth value of a summary is usually denoted as t or T. To avoid any confusion with t-norms, we adopt the letter v (considered also as validity).

The truth value of LS of the form Q R *entities in* $\mathbf{X}$ *are* P (we call it the structure with restriction) is computed as (Yager, 1982; Rasmussen and Yager, 1997)

$$v_{LSr}(\mathbf{X}) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_P(x_i) \wedge \sum_{i=1}^{n}\mu_R(x_i)}{\sum_{i=1}^{n}\mu_R(x_i)}\right), \tag{6}$$

where the membership function μ formalizes quantifier Q, restriction R, and predicate P of a considered data set $\mathbf{X}$. The restriction R focuses on a particular subdomain of attribute when evaluating relation among subdomains like *most of elderly patients have long stay*, where set *elderly* expresses R and *long stay* explains S. The convention $0/0 = 0$ is used in order to avoid undefined proportions (Rasmussen and Yager, 1997); this situation occurs when not a single record does meet R (and as a logical consequence, not a single record does simultaneously meet R and P) (Hudec *et al.*, 2018), where $\wedge$ stands for the conjunction. Theoretically, any t-norm function for conjunction can be applied (Klement *et al.*, 2000), but t-norms having downward reinforcement property unnaturally reduce the value of proportion. In the extreme case (drastic product t-norm), we get $T_{DP}(0.8, 0.9) = 0$. Considering this fact, the minimum t-norm can be the solution.

The truth value of a summary can be calculated by other approaches. The conjunction of Q *entities in* $\mathbf{X}$ *are* $P \wedge \hat{Q}$ *entities in* $\mathbf{X}$ *are* $\neg P$, where $\hat{Q}$ is an antonym of Q and $\neg P$

is a complement of P, is calculated by the method based on the Sugeno integral (Jain and Keller, 2015; Wilbik *et al.*, 2020). The same observation, as in the previous paragraph for the most suitable conjunctive function, holds here.

2.2. *Quality of Summaries*

The truth value is a significant measure, but it is not sufficient (Kacprzyk and Yager, 2001). Hirota and Pedrycz (1999) have introduced five features for measuring the quality of mined and aggregated information (not necessarily linguistically): validity (corresponds to truth value), novelty, usefulness, simplicity, and generality. Based on this observation, Wu *et al.* (2010) have proposed equations for calculating quality measures for LSs with restriction (see Eq. (6)) for transforming them into the IF-THEN rules. For instance, the degree of usefulness is computed as a minimum of truth value and coverage. Usually, quality measures focus on a single summary or a set of summaries to recognize the most representative ones (Bugarín *et al.*, 2015). Aggregation of several quality measures is examined in Hudec (2017). Kacprzyk and Strykowski (1999) have introduced quality measures: truth value, degree of fuzziness, degree of coverage, degree of appropriateness, and length of summary mainly related to the basic structure of LSs (see Eq. (5)). Recently, the quality measure conjunctively aggregating a summary and the summary consisting of antonym Q and negation of predicate P by the Sugeno integral for the summaries with restriction has been considered in Wilbik *et al.* (2020).

Another problem occurs when evaluating several subsets of data (in our case districts) by a single summarized sentence. The data distribution among regions varies and therefore it might cause skewed summaries, even though quality measures like data coverage, degree of fuzziness, or degree of focus do not report problems. However, not a single quality measure, to our best knowledge, is related to the quality of a summary evaluated over a hierarchical data, e.g. *the most of entities in a district have the low value of attribute A* calculated for each district. We need a quality measure for evaluating summaries on subsets of different sizes and compare the computed results. In the literature, theoretical works often illustrate achievements with smaller data sets for diverse summaries. The problem and proposed solution are discussed in the next section.

3. A New Quality Measure for a Single Summary Evaluated on Hierarchically Organized Subsets of Data

In this section, theoretical problems regarding the quality of the same LS (basic structure and structure with restriction) for all districts are evaluated and the new quality measure is proposed.

The foundation for the proposed approach is the theory of fuzzy sets introduced by Zadeh (1965), the theory of linguistic summarization proposed in Yager (1982) and Rasmussen and Yager (1997), the theory of aggregation functions summarized in Beliakov *et al.* (2007), and the aggregation functions of mixed behaviour proposed by De Baets and Mesiar (2002) and improved in Hudec *et al.* (2021). Hence, the methodology of our work is based on the key findings in these fields.

Table 1
Example of a truth value of summary on districts with different number of entities.

	District D1	District D2	District D3
Cardinality of set *young*	3	110	150
Cardinality of set *middle–aged*	3	120	140
Cardinality of set *old*	15	540	160
Truth value for summary *the most of patients are old* by (5)	0.7143	0.7013	0

3.1. *Basic Structure of LSs*

To illustrate the problem of a summary *Q entities in* $\mathbf{X}$ *are P*, let us evaluate the truth value of the sentence *the most of patients are old* for each district. The fuzzy set *old* is a linguistic term on a domain of attribute *age* (see e.g. Fig. 5). The quantifier *most of* is expressed by Eq. (4). In Table 1, there is an example of the numbers of patients in three districts, belonging to sets *young*, *middle-aged* and *old*. For the simplicity reason (which do not affect generality), all the patients belong to one of these sets with the degree equal to 1. In this case, $\sum_{i=1}^{n} \mu_P(x)$ (expressing the cardinality of a fuzzy set P) assigns a natural number.

District *D1* has a higher truth value than district *D2*, which means a slightly darker hue on a map. But, a higher concern should be focused on *D2*, instead of on *D1*. Thus, we should include the data distribution among districts to emphasize *D2* on a map. It has a significantly higher number of patients which should be reflected in the summary, while a lower number of patients should reduce the relevance (alarm) of a summary. Theoretically, not a single record in a district might be recorded, which leads to undefined operation $0/0$ in Eq. (5), i.e. $n = 0$. For this case, we adopt $0/0 = 0$.

The first (and simple) option is considering the proportion index p as a weight of the summary, i.e.

$$\overline{v_i}(\mathbf{X}) = p_i v_i(\mathbf{X}) \quad \text{for } i = 1, \ldots, n, \tag{7}$$

where n is the number of districts. The proportion can be expressed as

$$p_i = \frac{N_i}{\max\{N_1, \ldots, N_n\}}, \tag{8}$$

where N_i is the number of patients in district i, $N_i \leqslant \max\{N_1, \ldots, N_n\}$ causing $p_i \in [0, 1]$, $i = 1, \ldots, n$. If $\max\{N_1, \ldots, N_n\} = 0$, not a single case is recorded and therefore interpreting summaries on a map is irrelevant. For simplicity, we denoted $v_{LSb_i} = v_i$. Instead of the number of patients, the ratio of patients to the number of inhabitants in districts can be used. Firstly, it does not solve this problem. Secondly, many inhabitants might have a temporal address somewhere else. Assigning weights is a widely applied approach, but we should be careful. Discussions related to weights can be found in e.g. (Dujmović, 2018; Zadrożny *et al.*, 2008).

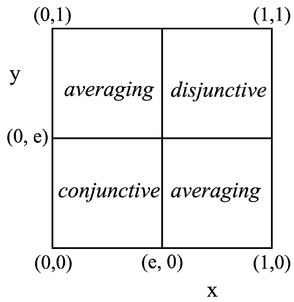

Fig. 2. The graphical interpretation of a uninorm function.

Next, $\overline{v_i}(\mathbf{X}) = v_i(\mathbf{X})$ holds only for a district where $N_i = \max\{N_1, \dots, N_n\}$. However, a high truth value for a district of a bit lower number of cases is attenuated, which is problematic.

Apparently, we should emphasize a summary for districts where both the truth value and proportion of data are high, and reduce the relevance of summaries when these two values are low. This observation leads to the aggregation by functions known as uninorms.

Uninorms generalize t-norms and t-conorms using the fact that these two classes of aggregation functions are defined by the same axioms of associativity, commutativity, monotonicity, and the presence of a neutral element. It means that uninorms consider neutral element e inside the unit interval (Beliakov *et al.*, 2007).

A uninorm is a bi-variate aggregation function $U : [0,1]^2 \to [0,1]$ which is associative, commutative, and has a neutral element $e \in]0,1[$. For $e \in \{0,1\}$ we have the limiting cases of t-conorm and t-norm. Next,

- $\forall (x,y) \in [0,e]^2 \quad U(x,y) = eT_u(\frac{x}{e}, \frac{y}{e})$ has a conjunctive behaviour;
- $\forall (x,y) \in [e,1]^2 \quad U(x,y) = e + (1-e)S_u(\frac{x-e}{1-e}, \frac{y-e}{1-e})$ has a disjunctive behaviour;
- $\forall (x,y) \in [0,e] \times [e,1] \cup [e,1] \times [0,e] \quad \min(x,y) \leqslant U(x,y) \leqslant \max(x,y)$ has an averaging behaviour,

where T_u stands for t-norm and S_u for t-conorm. This function is explained graphically in Fig. 2. When applying strict t-norm and strict t-conorm, we get the downward and upward reinforcement property, respectively (Beliakov *et al.*, 2007).

Representative uninorms are continuous everywhere except for the corners $(0,1)$ and $(1,0)$. For conjunctive uninorm holds $U(0,1) = 0$ (annihilator $a = 0$), whereas for disjunctive uninorm holds $U(0,1) = 1$ (annihilator $a = 1$). The former case might solve the problem with the quality of summaries. Due to commutativity, the same observation holds for $U(1,0)$. The presence of annihilator prevents uninorms from being strict on the whole unit square, i.e. they are strict on $]0,1[^2$.

An important family of parametrized representative uninorms is (Fodor *et al.*, 1997; Klement *et al.*, 1996):

$$U_\lambda(x,y) = \frac{\lambda xy}{\lambda xy + (1-x)(1-y)}, \qquad (x,y) \in [0,1]^2 / \{(0,1),(1,0)\}, \tag{9}$$

where $\lambda \in]0, \infty[$ and either $U_\lambda(0,1) = 0$ or $U_\lambda(0,1) = 1$, the neutral element is $e_\lambda = \frac{1}{1+\lambda}$. Taking $\lambda = 1$, we get the well-known $3 - \prod$ function (Yager and Rybalov, 1996)

$$U_\lambda(x, y) = \frac{xy}{xy + (1-x)(1-y)}, \tag{10}$$

where $e = 0.5$ and convention $0/0 = 0$ holds for the conjunctive uninorm.

This function can meet the needs for aggregating the truth value (consider $v_i = x$) and data proportion (consider $p_i = y$). When $p_i = 0$, the validity is also 0 (the validity of a summary on empty set is 0, when adopting $0/0 = 0$ in Eq. (5)). When $v_i = 0$, the solution should be 0, regardless of value p_i. These observations hold for the $3 - \prod$ function.

Moving back to the example in Table 1, we get $p_1 = 0.0273$, $p_2 = 1$, $p_3 = 0.5844$ by Eq. (8) and therefore we get by Eq. (7) $\overline{v_1} = 0.019$, $\overline{v_2} = 0.7013$ and $\overline{v_3} = 0$. The resulting relevance by Eq. (10) of summary for district *D1* is 0.0649 (averaging function), for district *D2* is 1 (disjunctive function), and for district *D3* is 0. District *D2* is emphasized because a truth value and the proportion of entities influencing a truth value are higher than 0.5. District *D1* is attenuated by the averaging behaviour of validity higher than 0.5 and the proportion lower than 0.5. District *D3* gets the value of 0 as the validity of summary is 0 (0 is an annihilator for the conjunctive behaviour).

On the other hand, due to discontinuity in the proximity of (0, 1) this function is unstable, especially for the imprecision of input data, i.e. $U(1, \epsilon) = 1$ for any $\epsilon > 0$ like 0.001. A simple solution is replacing 1 (for truth value and proportion) with 0.999 to get the average of 0.999 and 0.001. Anyway, we have searched for a better solution.

The next option is aggregating a truth value and data proportion by the ordinal sums, which are an extension for semigroups (Clifford, 1954) or for posets (Birkhoff, 1967). In the framework of fuzzy sets theory, they were considered to build new t-norms/t-conorms from the scaled versions of existing ones (Klement *et al.*, 2000). The ordinal sum of conjunctive and disjunctive functions has been proposed by De Baets and Mesiar (2002) as follows.

For an n-ary aggregation function $B : [0,1]^n \to [0,1]$ and $[a, b] \subset \mathbb{R}$, denote $B_{[a,b]}(\mathbf{x}) = a + (b-a) \cdot B(\frac{\mathbf{x}-a}{b-a})$. Then $B_{[a,b]}$ is an n-ary aggregation function on $[a, b]$. For $B_1, \ldots, B_k : [0,1]^n \to [0,1]$, $k \geqslant 2$, and $0 \leqslant a_0 < a_1 < \cdots < a_k = 1$. Let $A_i : [a_{i-1}, a_i]^n \to [a_{i-1}, a_i]$ be given by $A_i = (B_i)_{[a_{i-1}, a_i]}$. Then the ordinal sum $A : [0,1]^n \to [0,1]$, $A = (\langle a_{i-1}, a_i, A_i \rangle) | i = 1, \ldots, k$ given by De Baets and Mesiar (2002):

$$A(\mathbf{x}) = \sum_{i=1}^{k} \big(A_i \big(a_i \wedge (a_{i-1} \vee \mathbf{x}) \big) - a_{i-1} \big) \tag{11}$$

is an aggregation function on [0, 1]. If all $B_1, \ldots, B_k$ are t-norms (t-conorms, copulas, means) then also A is a t-norm (t-conorm, copula, mean) (De Baets and Mesiar, 2002).

Analogously, $A(\mathbf{x}) = \sum_{i=1}^{k} (a_i - a_{i-1}) \cdot B_i(1 \wedge (0 \vee \frac{\mathbf{x} - a_{i-1}}{a_i - a_{i-1}}))$. For our purposes, $n = k = 2$ is considered. Denoting $a_1 = a (a_0 = 0, a_2 = 1)$, we have two next forms of ordinal sums (Hudec *et al.*, 2021):

(i) $B_1, B_2 : [0,1]^2 \to [0,1]$,

$$A(x, y) = a \cdot B_1\left(1 \wedge \frac{x}{a}, 1 \wedge \frac{y}{a}\right) + (1-a) \cdot B_2\left(0 \vee \frac{x-a}{1-a}, 0 \vee \frac{y-a}{1-a}\right), \tag{12}$$

(ii) $A_1 : [0,a]^2 \to [0,a]$, $A_2 : [a,1]^2 \to [a,1]$,

$$A(x, y) = A_1(a \wedge x, a \wedge y) + A_2(a \vee x, a \vee y) - a. \tag{13}$$

Functions B_i cover subsquares of the unit square. In order to be inside the respective subsquares, conjunction between the attribute values (considering the separation point a) and 1, and disjunction between the values and 0 is applied. Observe that the function A covers the conjunctive part $A_1(x, y)$ when x and y are lower or equal to a, i.e. $a \wedge x = x$, and $a \wedge y = y$. Then, for A_2 we get $a \vee x = a$ and $a \vee y = a$. As a consequence $(a \vee a) - a = 0$. More details are in De Baets and Mesiar (2002), Hudec *et al.* (2021).

Then:

- if $(x, y) \in [0,a]^2$, $A(x, y) = a \cdot B_1(\frac{x}{a}, \frac{y}{a}) = A_1(x, y)$,
- if $(x, y) \in [a,1]^2$, $A(x, y) = a + (1-a) \cdot B_2(\frac{x-a}{1-a}, \frac{y-a}{1-a}) = A_2(x, y)$,
- if $(x, y) \in [0,a] \times [a,1]$, $A(x, y) = a \cdot B_1(\frac{x}{a}, 1) + (1-a) \cdot B_2(0, \frac{y-a}{1-a}) = A_1(x, a) + A_2(a, y) - a$,
- if $(x, y) \in [a,1] \times [0,a]$, $A(x, y) = a \cdot B_1(1, \frac{y}{a}) + (1-a) \cdot B_2(\frac{x-a}{1-a}, 0) = A_1(a, y) + A_2(x, a) - a$.

The next task is the suitable variation of conjunctive, disjunctive, and averaging functions in ordinal sums. In this work, we need upward reinforcement when both values are high, downward reinforcement when both are low and averaging behaviour when one measure is high and another is low. In addition, we need the stability in the ϵ neighbour of $(0, 1)$ and $(1, 0)$.

The option is a strict t-norm for the conjunctive part, strict t-conorm for the disjunctive part and a logically neutral averaging function (arithmetic mean), because we do not consider inclinations towards conjunctive or disjunctive areas provided by, e.g. geometric and quadratic mean, respectively.

The representative function of strict behaviour is product t-norm (expressed as $C_P(x, y) = xy$), whereas its dual t-conorm is a probabilistic sum (expressed as $D_P(x, y) = x + y - xy$).

The result for $A(0.5, 0.5)$, when $a = 0.5$ should be 0.5 for conjunctive, averaging and disjunctive function (see Fig. 3). In order to keep the expected value on this point, the product t-norm on $[0, 0.5]^2$ is expressed as (Hudec *et al.*, 2021):

$$C_P(x, y) = A_1(x, y) = 2xy. \tag{14}$$

Analogously, strict t-conorm on $[0.5, 1]^2$ is expressed as:

$$D_P(x, y) = A_2(x, y) = -1 + 2x + 2y - 2xy. \tag{15}$$

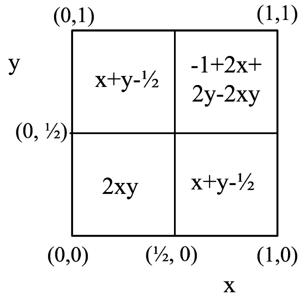

Fig. 3. The graphical interpretation of ordinal sums for product t-norm (14), probabilistic sum t-conorm (15) and arithmetic mean (16) (Hudec *et al.*, 2021).

Finally, the aggregation on the averaging part is expressed as:

$$AM(x, y) = A_1(x, 1/2) + A_2(1/2, y) - 1/2 = x + y - 1/2. \tag{16}$$

From the logic perspective, the arithmetic mean and its variants thereof (weighted arithmetic mean and the like) are the logically neutral averaging functions, with the *ORNESS* measure equal to 0.5. The other functions either incline towards conjunction (*ORNESS* < 0.5) or disjunction (*ORNESS* > 0.5) (Dujmović, 2018). Applying the other averaging functions will increase the complexity. This is the reason, why we applied arithmetic mean. However, the other averaging functions could be examined in the future work.

Considering again the example in Table 1 (consider $x = v$ – truth value and $y = p$ – proportion), the resulting quality of the summary ($\overline{v}$) for district *D1* is 0.2413 (an example of averaging behaviour), for district *D2* is 1 (an example of disjunctive behaviour), and for district *D3* is 0 (due to restriction put on summary). We get the expected results. Moreover, we get $A(1, 0.001) = 0.501$, which is an averaging behaviour. Next, when the proportion is 0, the truth value gets the same value. It is worth noting that, when a truth value is 0, the solution should be zero. This is not a problem, because quality measures should be activated when a truth value is greater than zero.

The downward and dual upward reinforcement behaviour is also a property of nilpotent t-norms and t-conorms, respectively. The representative functions are the Łukasiewicz t-norm and its dual Łukasiewicz t-conorm. In order to keep the expected value on edges of subinterval $[a, 1]^2$ when $a = 0.5$, the Łukasiewicz t-conorm on $[0.5, 1]^2$ is expressed as (Hudec *et al.*, 2021)

$$D_L(x, y) = A_2(x, y) = \min(1, x + y - 0.5). \tag{17}$$

When the truth value and proportion are higher than 0.75, the solution is equal to 1. Thus, we cannot distinguish between two summaries, for which the truth value and proportion are (0.75, 0.76) and (0.95, 1), respectively. Considering this fact and the need for providing continued hues on maps, the option is a strict t-norm and its dual t-conorm.

Table 2
Example of a ratio of included records and coverage.

Included records	i_c (Eq. (18))	coverage (Eq. (19))
175	0.0875	0.5377
320	0.1600	1
13	0.0065	0

3.2. *The Structure with Restriction*

In summary *Q R entities in* **X** *are P*, when, for instance, two of 10^6 entities satisfy R and only the same two entities satisfy P (Eq. (6)), the truth value gets value 1, but it is based on outliers (Hudec *et al.*, 2018). To overcome this problem, Wu *et al.* (2010) proposed a coverage measure that expresses the proportion of entities included in both R and P to avoid summaries on outliers. We refer this measure here. The ratio of included entities in a summary is

$$i_C = \frac{1}{n}\sum_{i=1}^{n} m_i, \tag{18}$$

where n is the number of records and $m_i = \begin{cases} 1 & \text{for } \mu_P(x_i) > 0 \wedge \mu_R(x_i) > 0, \\ 0 & \text{otherwise.} \end{cases}$

Since a summary of the structure (Eq. (6)) covers a subset of the entire database, i_c is considerably smaller than 1. Thus, the following function converts this ratio into the degree of sufficient coverage (Wu *et al.*, 2010)

$$C = f(i_c) = \begin{cases} 0 & \text{for } i_c \leqslant r_1, \\ 2(i_c - r_1)^2/(r_2 - r_1)^2 & \text{for } r_1 < i_c < (r_1 + r_2)/2, \\ 1 - 2(r_2 - i_c)^2/(r_2 - r_1)^2 & \text{for } (r_1 + r_2)/2 \leqslant i_c < r_2, \\ 1 & \text{for } i_c \geqslant r_2, \end{cases} \tag{19}$$

where the suggested values for parameters r_1 and r_2 depend on the length of the summary (the number of attributes in R and P).

To illustrate these calculations, let us have 2000 records, $r_1 = 0.02$ and $r_2 = 0.15$ (i.e. when 15% of data are included in the R and S parts, it is considered as a fully relevant coverage). Next, we have three summarized sentences S_1, S_2 and S_3 covering 175, 320 and 13 records, respectively. The ratio of included records and coverage are shown in Table 2. Summary S_2 fully covers the relevant subset of data, whereas summary S_3 should be excluded, even when its validity is equal to 1.

The method for calculating a truth value of conjunction of summary and its antonym based on the Sugeno integral (Jain and Keller, 2015) also solves this problem (Wilbik *et al.*, 2020).

The same problem as for a basic structure of a summary holds here. Even though quality measures filter summaries on outliers, the distinction among subsets of different sizes should be reflected on the map. Districts, where the number of patients is higher, should be emphasized when the truth value of a summary and data coverage are high. In addition, we have a truth value of a summary, data coverage, and proportion. Hence, we should aggregate these three measures.

The truth value and coverage are measures for evaluating different summaries on the same data set. As both should be satisfied, we aggregate them by t-norm function (Hudec, 2017). In the next step, we apply ordinal sums (see Eqs. (14), (15), (16)).

4. Interpreting Linguistic Summaries on Maps

A single statistical figure can be interpreted by hues of a selected colour reflecting the values for each district (or variants thereof). These hues can be continuous coverage from the smallest to the highest value of the considered statistical figure. The next option is dividing a domain of the possible values into several categories which could be equi-length, equi-depth and equi-log (where log stands for logarithm) (Aggarwal, 2015). Consequently, each category gets its unique colour. Next, bar and pie charts are suitable for explaining several figures for each region (when figures cannot be aggregated into single figure). Usually, charts require training and experience to be interpreted quickly and accurately (Reiter, 2017). A possible solution is interpreting the quality of a LS on maps.

The validity of a LS gets value from the unit interval. Thus, we can straightforwardly convert this value into the hues by the chosen colour. The next option is applying one colour for value 0, applying another colour (quite different) for value 1 and for values in]0, 1[applying the hues of the third colour. Observe that this interpretation is analogous to classification into three overlapping classes (*yes*, *no*, and *maybe* with indicating inclination towards classes *yes* and *no*) (Hudec *et al.*, 2021).

Quantifier *most of* (see Eq. (4)), and its variants thereof, for the proportions greater than or equal to 0.8 assigns truth value equal to 1. It is a convenient way to explain storytelling like *the most of young customers buy groceries in the early evening, the most of middle-aged customers buy groceries around the noon, the most of old customers buy groceries in the morning*. In our task, we explore the same summary among the disjoint subsets of data. When we adopt the quantifier (Eq. (4)) for visualizing districts by the sentence *the most of patients are young*, for instance, then the user would appreciate a difference between the proportions of 0.81 and 0.99. Next, proportions lower than 0.5 should be also evaluated. This problem can be solved by modifying quantifier (Eq. (3)) to be strictly linear (or non-linear) function as

$$\mu_Q(y) = \max\left(0, \min\left(1, \frac{y - 0.3}{0.7}\right)\right). \tag{20}$$

All districts having a proportion lower than or equal to 0.3 get a truth value of 0. It is an acceptable solution. However, the next question is enveloping low proportions. Such proportions can be covered by the quantifier *few* to emphasize districts with a low proportion of entities satisfying a predicate. Interpreting two sentences on a map might be confusing. Thus, we propose the quantifier *significant proportion* as:

$$\mu_Q(y) = y, \quad \text{for } y \in [0, 1], \tag{21}$$

which meets the requirements $\mu_Q(0) = 0$ and $\mu_Q(1) = 1$. Its antonym a *insignificant proportion* is expressed as:

$$\mu_Q(y) = 1 - y, \quad \text{for } y \in [0, 1]. \tag{22}$$

In this way, we get a strictly increased (and decreased, respectively) function for expressing the proportion of entities satisfying a summarized sentence and therefore distinguishing districts by hues of the selected colour. Generally, fuzzy relative quantifiers can be formalized by non-linear functions. In the case of non-linear functions, the users have to specify the shapes, which is not a simple task for domain experts, the case in the medical domain (Holzinger *et al.*, 2017). Hence, we adopted the linear ones due to their simplicity for the end users.

The next section explains the technical background for interpreting summaries on maps and illustrates the developed model on real data.

5. Experiments on Data

At the beginning of the pandemic, the cases were rare. Thus, the approach based on LSs was not suitable for interpreting aggregated data. Within one year of the pandemic situation, a significant number of cases were recorded. It opens the space for the application of linguistic summaries.

The application is realized for all 79 districts of the Slovak Republic. The country has approximately 5 400 000 inhabitants. The data source consists of 13 967 records for all 79 districts collected in 2021 in one of the three health insurance companies in Slovak Republic, which takes care of the health of about 30% of the population of Slovak Republic. The number of cases has significantly increased to the end of 2020. The culmination was registered in early spring of 2021. The data are in a matrix form, which is usual for data analysis. The personal data were fully anonymized before transferring from the health insurance company to this model. The only data related to patients are age calculated from the year of birth and the district of the patient's permanent residence.

To illustrate the proposed approach, we created a simplified web application. The structure of this application was created using Hyper Text Markup Language (HTML), whereas the dynamic content was added using Hypertext Preprocessor (PHP). We used a MySQL database to store and query data. After executing a flexible quantified query over the data stored in the database, the results are saved to a JSON file, so that they can be used for displaying on a map. To interpret the query results on a map by hues of the selected colour, the JavaScript library Leaflet has been adopted. It is a simple open-source JavaScript library for interactive maps (more on https://leafletjs.com/).

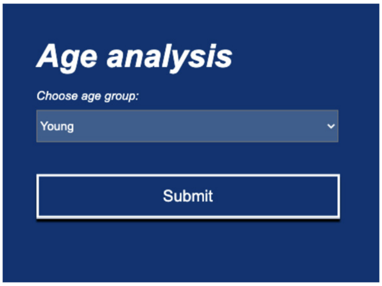

Fig. 4. The interface for a basic structure of a LS for attribute *age*.

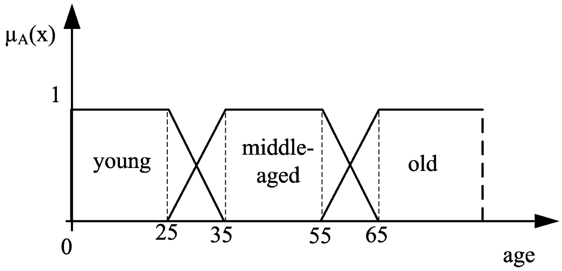

Fig. 5. The LV for attribute *age*.

To display the polygons of districts on a map, their exact coordinates are required. These coordinates should be in a format usable for programming. For this reason, we used the GeoJSON file, which is a format for encoding a variety of geographic data structures (more on https://geojson.org/). Next, the calculations of the colour hue of the district using the JavaScript language have been implemented. The last step was linking the query results with the polygons of the districts on the map using the district name because the GeoJSON file contains names of the district as an identifier.

After launching the application, the user has the option to choose between age analysis, age analysis of death cases, and relations between the length of hospitalization and age. Age analysis and age analysis of death cases are basic linguistic summaries, whereas the analysis of length of hospitalization and age relationship is the linguistic summary with restriction. A user also chooses which linguistic terms he/she wants to analyse by selecting them in a simple form shown in Fig. 4. Linguistic variable for attribute *age* consisting of three terms is shown in Fig. 5.

The solution for the summary is visualized on the map of the Slovak Republic in Fig. 6. The solution is obtained by aggregating the truth value (Eq. (5)) with the data proportion by the ordinal sums (Eqs. (14), (15), (16)). To illustrate the problem of applying only the truth value, we can observe the interpretation in Fig. 7, where the Myjava district is indicated as a critical one, but the total number of patients in this district is very low in comparison to the Košice district. It is the same problem as the problem illustrated in Table 1. The LV *age* reflects a usual consideration of age. Thus, the degree of fuzziness is correct. When using an attribute like the number of sold items, then *low*, *medium*, and

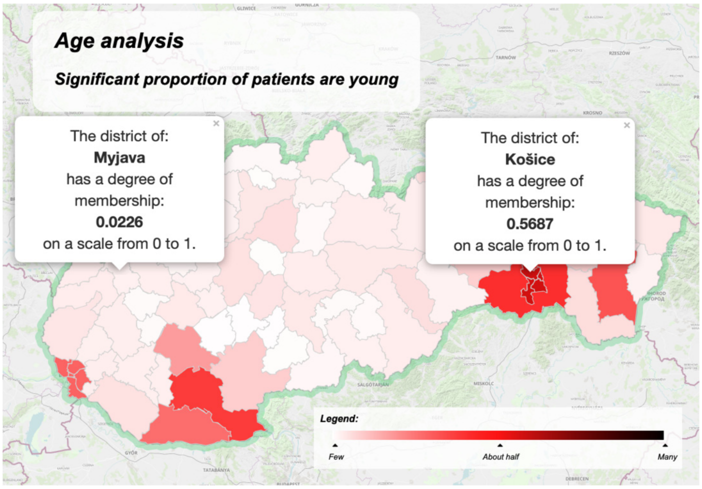

Fig. 6. Interpreting the basic structure of the summary *significant proportion of patients are young* by the proposed quality measure.

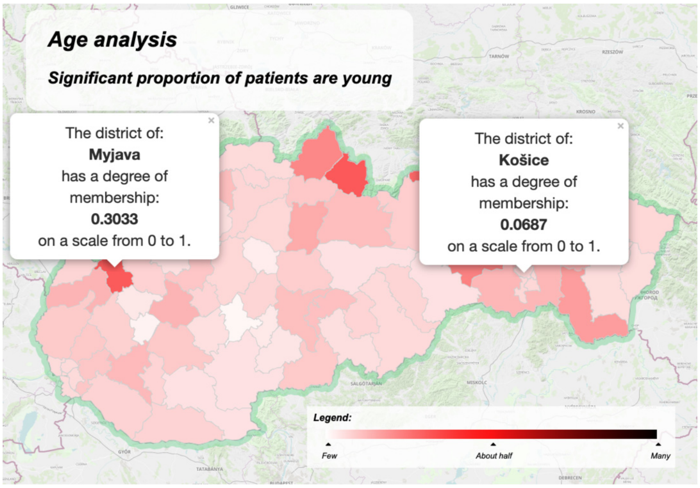

Fig. 7. Interpreting the basic structure of the summary *significant proportion of patients are young* considering only truth value.

high significantly vary for diverse products and therefore the measure of fuzziness can be adopted to increase the quality of the summary.

The evaluated question of the structure with restriction (6) is the following: *the significant proportion of old patients has a long stay in hospitals*. The linguistic term *long* of the variable *stay* has parameters $x_3 = 12.5$ and $x_4 = 15$ (see Fig. 1). The user interface depicted in Fig. 8 is simple, but intuitive where the user selects age category and length of hospitalization, both expressed linguistically. The solution for this summary by the pro-

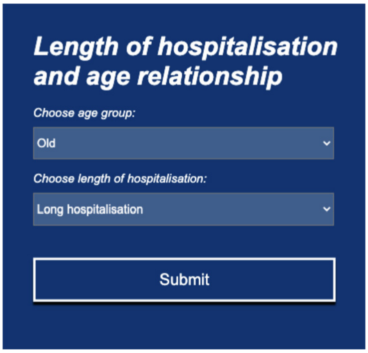

Fig. 8. The interface for a structure with restriction of a LS for attributes *age* and *length of hospitalisation*.

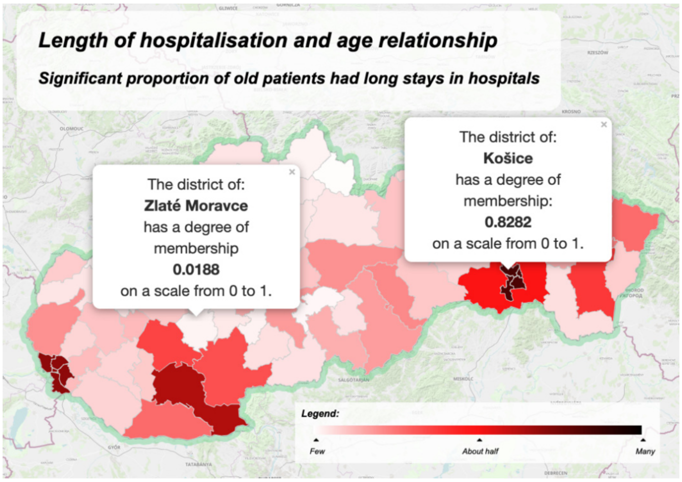

Fig. 9. Interpreting the summary with restriction *significant proportion of old patients has a long stay in hospitals* by the proposed quality measure.

posed quality measure is visualized in Fig. 9, whereas the solution considering only the truth value is shown in Fig. 10.

The next option is the comparison of developments between the years 2020 and 2021. The summary *significant proportion of old patients have a long stay in hospitals* for 2021 is shown in Fig. 9, whereas the same summary for the 2020 year is shown in Fig. 11. We can see not only where the waves had a stronger impact on elderly people, but also how the waves had moved among the districts.

6. Discussion

This approach proposes a novel way for interpreting linguistically summarized sentences on maps. The question is posed linguistically by selecting attributes and adjectives from the list of linguistic terms.

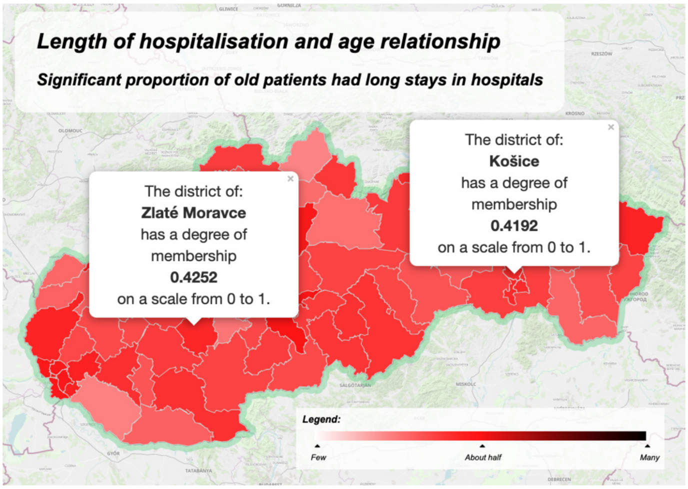

Fig. 10. Interpreting the summary with restriction *significant proportion of old patients has a long stay in hospitals* considering only the truth value.

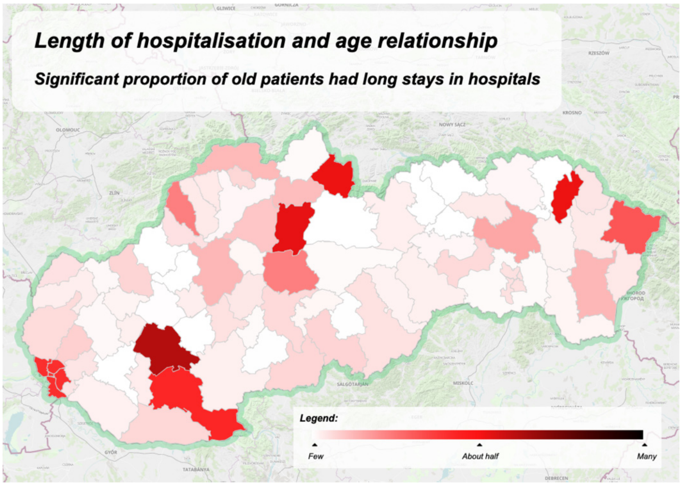

Fig. 11. Interpreting the summary with restriction *significant proportion of old patients has a long stay in hospitals* for year 2020.

Linguistic terms are vague, but very effective. Here, "vague" means non-sharp boundaries expressed by fuzzy sets, whereas "effective" means that we distinguish elements by the intensity of belonging to a set without adding further properties (Radojević, 2008).

The solution is realized on the data from one health insurance company (DÔVERA zdravotná poisťovňa, a.s.). This solution provides an insight into the distributions of patients for this company. Currently, three health insurance companies covering health services guaranteed by the government operate in Slovak Republic. To get the full overview of the situation regarding the COVID-19 cases for governmental organizations, journalists, researchers, and the general public, data from all health insurance companies should be merged. This integration does not affect the mathematical and programming background as one patient is a client of only one health insurance company. It is worth noting that this solution is not restricted to COVID-19 cases. It can be extended to monitor other illnesses. Moreover, this solution can be adapted to monitor the development in other areas like pollution among districts and socio-economic aspects, which might augment (official statistics) data dissemination (Hudec *et al.*, 2018) by interpreting summaries on maps.

This concept can easily be applied to any human language. Adjectives expressing fuzzy sets like *high*, *long* and *old*, and quantifiers such as *significant amount of*, *most of* and *almost all* are always expressed by increasing functions (linear or non–linear), regardless of their translation to other languages and examined concepts.

The quality measures for linguistic summaries and parameters of fuzzy sets are computed from the data. While this is a convenient solution, we should also consider experts at a health insurance company (or any other experts). For such users it might be useful to adjust parameters to meet particular requirements.

The further important activity of research is to develop advanced designs of easy-to-use application interfaces for diverse categories of users (as experts in the field and the general public). Developing and testing the interfaces can be realized within the health insurance companies and requires cooperation between health insurance experts, web developers, especially web designers, and data dissemination experts.

The proposed approach explains summaries from the data (by interpreting them on a map), not the data itself. Generally speaking, the data disclosure in summarization would not be a problem; however, care should be taken when summarizing from small data sets. The decision which data might be available to realize summaries and interpret them on maps should meet regulations and other relevant rules when disseminating them to public.

Considering the interpretation of summaries on maps, the future task will be focused on evaluating the proposed quality measure and quantifiers on diverse data sources and their potential in various domain areas. For example, creating the business intelligence dashboards that respect specifics of strategic, tactic, and operational needs in business (Vaisman and Zimányi, 2014).

7. Conclusion

The interpretation of linguistically summarized sentences on maps provides a quick overview of the developments in districts. In order to contribute to this field, we raised research questions of calculating the quality of summaries for each district and interpreting them on maps.

The existing quality measures are focused on a single summary or to find the most suitable summary from the set of summaries. In this work, we proposed a new quality measure for evaluating the same summary on different subsets of hierarchical data (i.e. the number of patients by districts). This quality measure considers data proportion among districts. The answer to our research questions is that we can express the relevance of a summary on a map by aggregating proportions of data among districts and the truth value by the recently developed ordinal sums of conjunctive and disjunctive functions. More precisely, only strict conjunctive and disjunctive functions are suitable. Consequently, the result (assigned value from the unit interval) is interpreted on a map.

The potential of linguistic summaries in interpreting and disseminating summarized information on maps is demonstrated on real-world data regarding COVID-19 cases. In order to reduce the burden on users for using the quality measure and visualization of the short-quantified sentences on maps, we have developed an interface for selecting attributes and their adjectives. It brings a user-friendly environment for visualization without deeper knowledge of the data and the computing process.

Further, our research has documented perspectives for the application of the proposed method in other health insurance tasks and beyond. The next task will be focused on evaluating the proposed quality measure and quantifiers on diverse data sources and for different tasks (e.g. mapping civilization diseases, environmental problems, distribution of different types of enterprises in the regions, the poverty rate in the regions). Finally, we underline that the proposed approach should be considered as a complementary data interpretation to the established practice of interpreting statistical figures on maps.

Acknowledgments

The authors would like to thank the health insurance company DÔVERA zdravotná poisťovňa, a. s., Bratislava, Slovak Republic for advice regarding the research topic and provided anonymized data.

Funding

This work was partially supported by the Ministry of Education, Youth and Sports of the Czech Republic under Grant SGS No. SP2021/86; and partially supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under Grant 025EU-4/2021.

References

Aggarwal, C.C. (2015). *Data Mining: The Textbook*. Springer, Cham.

Alonso, J.M., Castiello, C., Magdalena, L., Mencer, C. (2021). *Explainable Fuzzy Systems*. Springer, Cham.

Beliakov, G., Pradera, A., Calvo, T. (2007). *Aggregation Functions: A Guide for Practitioners*. Springer, Berlin Heidelberg.

Birkhoff, G. (1967). *Lattice Theory*. *Colloqium Publications*, Vol. XXV, 3rd ed. American Mathematical Society, Providence.

Boran, F.E., Akay, D., Yager, R.R. (2016). An overview of methods for linguistic summarization with fuzzy sets. *International Journal of General Systems*, 61, 356–377.

Bugarín, A., Marín, N., Sánchez, D., Trivino, G. (2015). Aspects of quality evaluation in linguistic descriptions of data. In: *Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015)*, Istanbul, Turkey, August 2–5.

Clifford, A.H. (1954). Naturally totally ordered commutative semigroups. *American Journal of Mathematics*, 76, 631–646.

De Baets, B., Mesiar, R. (2002). Ordinal sums of aggregation operators. In: *Technologies for Constructing Intelligent Systems*, Vol. 2. Springer, Berlin Heidelberg, pp. 137–147.

Dujmović, J. (2018). *Soft Computing Evaluation Logic: The LSP Decision Method and Its Applications*. John Wiley and Sons, Hoboken.

Eurostat (2003). *Regions Nomenclature of Territorial Units for Statistics*. Office for Official Publications of the European Communities, Luxembourg.

Fodor, J., Yager, R.R., Rybalov, A. (1997). Structure of uninorms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5, 411–427.

Hirota, K., Pedrycz, W. (1999). Fuzzy computing for data mining. *Proceedings of IEEE*, 87, 1575–1600.

Holzinger, A., Malle, B., Kieseberg, P., Roth, P.M., Müller, H., Reihs, R., Zatloukal, K. (2017). Machine learning and knowledge extraction in digital pathology needs an integrative approach. In: Holzinger, A., Goebel, R., Ferri, M., Palade, V. (Eds.), *Towards Integrative Machine Learning and Knowledge Extraction*. Springer, Cham, pp. 13–50.

Hudec, M. (2017). Merging validity and coverage for measuring quality of data summaries. In: Kulczycki, P., Kóczy, L.T., Mesiar, R., Kacprzyk, J. (Eds.), *Information Technology and Computational Physics*. Springer, Cham, pp. 71–85.

Hudec, M., Bednárová, E., Holzinger, A. (2018). Augmenting statistical data dissemination by short quantified sentences of natural language. *Journal of Official Statistics*, 34(4), 981–1010.

Hudec, M., Mináriková, E., Mesiar, R., Saranti, A., Holzinger, A. (2021). Classification by ordinal sums of conjunctive and disjunctive functions for explainable AI and interpretable machine learning solutions. *Knowledge–Based Systems*, 220, 106916.

Jain, A., Keller, J.M. (2015). On the computation of semantically ordered truth values of linguistic protoform summaries. In: *Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015)*, Istanbul, Turkey, August 2–5.

Kacprzyk, J., Strykowski, P. (1999). Linguistic data summaries for intelligent decision support. In: *Proceedings of the fourth European Workshop on Fuzzy Decision Analysis and Recognition Technology for Management, Planning and Optimization (EFDAN 1999)*, Dortmund, Germany, June 14–15.

Kacprzyk, J., Yager, R.R. (2001). Linguistic summaries of data using fuzzy logic. *International Journal of General Systems*, 30(2), 133–154.

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

Kacprzyk, J., Zadrożny, S. (2010). Modern data-driven decision support systems: the role of computing with words and computational linguistics. *International Journal of General Systems*, 39(4), 133–154.

Kacprzyk, J., Wilbik, A., Zadrożny, S. (2006). Linguistic summarization of trends: a fuzzy logic based approach. In: *Proceedings of the 11th Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU 2006)*, Paris, France, July 2–7.

Klement, E.P., Mesiar, R., Pap, E. (1996). On the relationship of associative compensatory operators to triangular norms and connorms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 4, 129–144.

Klement, E.P., Mesiar, R., Pap, E. (2000). *Triangular Norms*. Kluwer, Dordrecht.

Lesot, M.-J., Moyse, G., Bouchon-Meunier, B. (2016). Interpretability of fuzzy linguistic summaries. *Fuzzy Sets and Systems*, 292, 307–317.

Liétard, L. (2006). A new definition for linguistic summaries of data. In: *Proceedings of the 2008 IEEE International Conference on Fuzzy Systems*, Hong Kong, China, June 1–6.

Paliulionis, V. (2000). Intelligent GIS: architectural issues and implementation methods. *Informatica*, 11, 269–280.

Radojević, D. (2008). Interpolative realization of Boolean algebra as a consistent frame for gradation and/or fuzziness. In: Nikravesh, M., Kacprzyk, J., Zadeh, L.A. (Eds.), *Forging New Frontiers: Fuzzy Pioneers II, Studies in Fuzziness and Soft Computing*. Springer-Verlag, Berlin Heidelberg, pp. 295–318.

Rasmussen, D., Yager, R.R. (1997). Summary SQL – a fuzzy tool for data mining. *Intelligent Data Analysis*, 1, 49–58.

Reiter, E. (2017). Non-experts struggle with information graphics. https://ehudreiter.com/2017/10/02/non-experts-struggle-graphs/. Accessed 4 February 2021.

Ruspini, E.H. (1969). A new approach to clustering. *Information and Control*, 15(1), 22–32.

Vaisman, A., Zimányi, E. (2014). *Data Warehouse Systems – Design and Implementation*. Springer, Berlin Heidelberg.

Wilbik, A., Havens, T.C., Wilkin, T. (2020). On a paradox of extended linguistic summaries. In: *Proceedings of the 2020 IEEE International Conference on Fuzzy Systems*, Glasgow, UK, July 19–24.

Wu, D., Mendel, J.M., J., J. (2010). Linguistic summarization using if-then rules. In: *Proceedings of the 2010 IEEE International Conference on Fuzzy Systems*, Barcelona, Spain, July 18–23.

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

Yager, R.R., Rybalov, A. (1996). Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80, 111–120.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8, 338–353.

Zadeh, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning: Part I. *Information Sciences*, 8, 199–249.

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

Zadrożny, S., De Tré, G., De Caluwe, R., Kacprzyk, J. (2008). An overview of fuzzy approaches to flexible database querying. In: Galindo, J. (Ed.), *Handbook of Research on Fuzzy Information Processing in Databases*. Information Science Reference, Hershey, pp. 34–54.

M. Hudec is an associate professor at the University of Economics in Bratislava, Faculty of Economic Informatics (Slovak Republic) and VSB – Technical University of Ostrava, Faculty of Economics (Czech Republic). He received the PhD degree from the University of Belgrade (Serbia). His work is mainly focused on fuzzy logic, knowledge discovery and business intelligence. He is a member of program committees of several international conferences and serves as an associate editor in *Applied Soft Computing*. He has published more than 50 articles including a monograph in Springer.

K. Malovcová received a master's degree from the University of Economics in Bratislava, Faculty of Economic Informatics (Slovak Republic) in June 2021. In her final thesis she dealt with the flexible calculation of linguistic summaries and interpretation of results on a map. She works as a junior analyst at the health insurance company DÔVERA zdravotná poisťovňa, a.s., Slovak Republic.

R. Trumic holds a master in business administration from the University of Applied Sciences Munich, Germany. Currently, he is a PhD student at the Faculty of Economics, VSB – Technical University of Ostrava, Czech Republic. He has been working in supply chain management for BMW in Munich for many years and his research is mainly related to cost avoidance in change management.

E. Rakovská received her PhD degree in applied informatics at the University of Economics in Bratislava (Slovak Republic) in 2010. She received a master's degree in mathematics at Comenius University in 1986. She worked as a programmer and IT developer in various companies, including a hospital. She is an assistant professor at the University of Economics in Bratislava, Faculty of Economic Informatics since 2001. Her research interests are oriented toward applied informatics and knowledge engineering. The main impact is on the soft computing application and expert systems.