Type-2 Fuzzy Summarization of Data: An Improved News Generating

Adam Niewiadomski

Institute of Computer Science, Technical University of Lodz
ul. Wólczańska 215, 90-924 Łódź, Poland
aniewiadomski@ics.p.lodz.pl

Abstract. The paper introduces an improved method of intelligent summarization of large datasets. Previously, the author's solution for automated generating of textual news and comments, based on the standard Yager's method and ordinary fuzzy sets, has been published in [1]. In this paper, a type-2-fuzzy-set-based extension of the concept can be now introduced. Type-2 membership functions are originally applied to build new summarization methods. The approach generalizes the previous methods which are based on traditional fuzzy sets. Moreover, new quality measures of summaries are proposed and used in selecting the optimal and the most specific summaries as the components of textual news. Finally, the method is implemented and evaluated.

1 Motivation and Problem Study

The problem of distilling useful and ready-to-use knowledge from huge amounts of unstructured and dispersed data, is very present now. The original concept of *a linguistic summary of a database* introduced by R. R. Yager in 1982 [2] appeared a simple and effective methods. Linguistic summaries are natural language sentences that approximately but clearly describe properties of objects, e.g. *About 100 of my students are excellent programmers*, where *students* is the subject of summary, and *about 100* and *excellent programmers* are pronouncements of amount and property, respectively, both handled by fuzzy logic [3,4].

The gist of the paper is to enhance the Yager method with the use of *Type-2 Fuzzy Sets* [5]. They extend the Zadeh idea, and enable representing imprecise information via type-2 membership functions which are fuzzy-valued functions. Since traditional membership functions may appear inconsistent as they represent imperfect information via precise and crisp numbers, the use of type-2 membership functions as models of vague quantities and features needs to be discussed. Some research on type-2 fuzzy sets in linguistic data summarization have already been made by the author in [6,7,8,9].

The main motivation to generalize the Yager approach is that membership degrees of properties or phenomena under many circumstances may be inexpressible in terms of crisp values. Type-1 membership functions are frequently constructed based on preferences of one expert. However, it may look arbitrary, since it seems more natural when two or more opinions are given to illustrate

M. Kryszkiewicz et al. (Eds.): RSEISP 2007, LNAI 4585, pp. 241–250, 2007.

e.g. a linguistic term, to model it as objectively as possible. Traditional fuzzy sets dispose no methods of handling these, usually different, opinions. The average or median of several membership degrees keep no information about those natural differences. For instance, the question *What is the compatibility level of the* $36.5°C$ *with "temperature of a healthy human body"?* can be answered 0.5, 1.0, 1.0 by three doctors, respectively, but the average, 0.866, does not show that one of them remains unconvinced.

Extending real (type-1) membership levels to fuzzy-valued (type-2) provides additional computational tools – secondary membership degrees. They may be interpreted as *possibility levels* that primary degrees describe memberships appropriately, but from the point of view of the linguistic summarization, interpreting them as *weights* [5] is practicable. Thanks to it, different expert opinions on a membership degree may be described by "confidence levels" which express e.g. expert's experience. See the example on the temperature (above): the proposed compatibility values may be presented as $(0.5, 0.2), (1.0, 1.0), (1.0, 0.9)$ which says that the first expert is much less experienced than the others and this information is stored in the resulting fuzzy set. This set may be – but *need not* to be – defuzzified or averaged. The goal is to use different types of fuzzy sets when generating summaries, and to maintain the understandable semantics of results (real degrees of truth and other quality measures) proposed by Yager. Thus, we present a general method of summarization in which many types of fuzzy sets may be applied, and the differences among them are hidden for an end-user.

2 Information Representation Via Type-2 Fuzzy Sets

2.1 Basic Definitions

The idea of a type-2 fuzzy set extends an ordinary membership function to a *type-2 membership function*. This is a family of type-1 sets in $[0, 1]$ assigned to elements of a universe of discourse. A *type-2 fuzzy set* $\tilde{A}$ in $\mathcal{X}$ is defined $\tilde{A} = \int_{\mathcal{X}} \mu_{\tilde{A}}(x)/x$ and $\mu_{\tilde{A}}: \mathcal{X} \to \mathcal{F}([0, 1])$ is the type-2 membership function, such that $\mu_{\tilde{A}}(x) = \int_{u \in J_x} \mu_x(u)/u$, $J_x \subseteq [0, 1]$. Each u has its own membership degree assigned. Moreover, many u's can be assigned to a given x, and each has its separated *secondary* membership degree $\mu_x(u)$. For a fixed x', $\mu_{x'}$ is the membership function for the type-1 set which expresses the membership of x' to $\tilde{A}$, i.e. for the $\mu_{\tilde{A}}(x')$ value. Secondary degrees may be viewed as *weights* or as *possibility levels*, cf. [5].

The set-theoretical operations on type-2 sets are extensions of the analogous ones in other fuzzy set theories. Let $\tilde{A}$, $\tilde{B}$ be type-2 sets in $\mathcal{X}$. Let t_1, t_2 be t-norms. The intersection of $\tilde{A}$ and $\tilde{B}$ is the type-2 set $\tilde{A} \cap \tilde{B}$, the membership function of which is defined in terms of *the meet* operation:

$$\mu_{\tilde{A}\cap\tilde{B}}(x) = \mu_{\tilde{A}}(x) \sqcap \mu_{\tilde{B}}(x) = \int_{u_{\tilde{A}}} \int_{u_{\tilde{B}}} (\mu_x(u_{\tilde{A}})\ t_1\ \mu_x(u_{\tilde{B}}))/(u_{\tilde{A}}\ t_2\ u_{\tilde{B}}) \tag{1}$$

where $u_{\tilde{A}}$, $u_{\tilde{B}}$ – primary membership degrees of x in $\tilde{A}$, $\tilde{B}$, respectively; $\mu_x(u_{\tilde{A}})$, $\mu_x(u_{\tilde{B}})$ – secondary degrees of x in $\tilde{A}$, $\tilde{B}$, respectively. Eq. (1) is applied as a model for the AND connective that combines single summarizers, see Sec. 3.

The concept of *embedded fuzzy set* appears useful in defining other concepts. *an embedded type-1 set* A_λ for a type-2 fuzzy set $\tilde{A}$ in $\mathcal{X}$, is defined. Let $\forall_{x\in\mathcal{X}}\ \lambda_x \in J_x \subseteq [0,1]$. The membership function for A_λ is given as $\mu_{A_\lambda}(x) = \lambda_x$,

2.2 Cardinality, Support, and Degree of Imprecision of Type-2 Sets

Cardinality of a crisp set A' in $\mathcal{X}$ is the sum of the $\xi_{A'}$ characteristic function values $card(A') = \sum_{x\in\mathcal{X}} \xi_{A'}(x)$. The cardinality of a type-1 set A in $\mathcal{X}$ [10]

$$card(A) = \sum\nolimits_{x\in\mathcal{X}} \mu_A(x) \tag{2}$$

The cardinality of a type-2 set, *non-fuzzy sigma count*, assumes that membership of x in $\tilde{A}$ in $\mathcal{X}$ is a fuzzy number. Hence nfσ-count($\tilde{A}$) is defined:

$$\text{nf}\sigma\text{-count}(\tilde{A}) = \sum\nolimits_{x\in\mathcal{X}} \max\{u \in J_x \colon \mu_x(u) = 1\} \tag{3}$$

The given definition is a generalization of the analogous definition for an ordinary fuzzy set, given by de Luca and Termini [10].

The support of a type-1 set is defined as

$$supp(A) =_{df} \{x \in \mathcal{X} \colon \mu_A(x) > 0\} \tag{4}$$

and is applied to measure the goodness of summaries. We propose to extend it to *the fuzzy support* – a set of type-1 associated with a given type-2 set.

Definition 1. *Let $\widetilde{A}$ be a type-2 set in $\mathcal{X}$.* The fuzzy support of $\widetilde{A}$ *is the type-1 set $supp(\widetilde{A}) =_{df} \{\langle x, \mu_{supp(\widetilde{A})}\rangle \colon x \in \mathcal{X}\}$ where*

$$\mu_{supp(\widetilde{A})}(x) = \sup_{u\in J_x\setminus\{0\}} \mu_x(u) \tag{5}$$

Proposition 1. *For each type-1 fuzzy set A, $\mu_{supp(A)}(x) = \xi_{A_0}(x)$*

Proof. Let A be a type-1 fuzzy set in $\mathcal{X}$. Hence, each its element has only one primary membership value assigned, $u(x)$, and $\forall_{x\in\mathcal{X}}\ \mu_x(u) = 1$, so the supremum in (5) can be omitted. Thus, $supp(A)$ – the zero-cut of A, is a crisp set.

Definition 2. *Let $\widetilde{A}$ be a type-2 set in $\mathcal{X}$. The degree of fuzziness of $\widetilde{A}$ is defined:*

$$in(\widetilde{A}) =_{df} card(supp(\widetilde{A}))/card(\mathcal{X}) \tag{6}$$

The definition extends the concept for type-1 sets, and is applied to determine quality indices of type-2 summaries in Sec. 3.3.

2.3 Type-1 Fuzzy Quantification of Type-2 Fuzzy Propositions

The canonical forms of linguistically quantified propositions are defined in [4]. We originally generalize them with type-2 sets as models of S_1, S_2

Definition 3. *Let $\widetilde{S}_1$, $\widetilde{S}_2$ be type-2 sets representing linguistic propositions, and Q – a type-1 fuzzy quantifier. The formulae*

$$Q\ x\text{'s are } \widetilde{S_1} \tag{7}$$

$$Q\ x\text{'s being } \widetilde{S_2} \text{ are } \widetilde{S_1} \tag{8}$$

are the first *(Q^I) and* the second canonical form (Q^{II}) *of the linguistically quantified proposition. Degrees of truth of (7) and (8) are assessed as*

$$T(\ Q\ x\text{'s are } \widetilde{S_1}\ ) = \mu_Q(card(\widetilde{S_1})/M) \tag{9}$$

where $card(\widetilde{S_1})$ is a real number, see (3), $M = card(\mathcal{X})$ if Q is relative, or $M = 1$ if Q is absolute, and

$$T(\ Q\ x\text{'s being } \widetilde{S_2} \text{ are } \widetilde{S_1}) = \mu_Q(card(\widetilde{S_1} \cap \widetilde{S_2})/card(\widetilde{S_2})) \tag{10}$$

where $\widetilde{S_1} \cap \widetilde{S_2}$ is given in (1).

Examples for Q^I, Q^{II}, are MANY *students are intelligent* and MANY *of young students are intelligent*, respectively, in which MANY=Q, *intelligent*=S_1, and *young*=S_2. Similarly to the propositions represented by type-1 sets, only relative quantification is possible in (8).

3 Type-2 Linguistic Summaries of Data

The section introduces the linguistic data summarization algorithms innovated by the use of type-2 fuzzy logic. In particular, we are interested in the Q P are/have $\widetilde{S}$ $[T]$, form of summary, in which $\widetilde{S}$ is a summarizer represented by a type-2 fuzzy set, and Q, P, T are interpreted as in type-1 summaries.

3.1 Type-2 Summaries in the First Canonical Form

We introduce the type-2 summaries based on Q^I, see (7). The goal is to find a quality index for a given summary in the form of Q P are/have $\widetilde{S}$. We assume here, that Q is represented by a type-1 fuzzy set and the cardinality of $\widetilde{S}$ is computed via (3). The degree of truth of such a summary is a real number

$$T\left(Q\ P\ \text{ are/have } \widetilde{S}\right) = \mu_Q(\text{nf}\sigma\text{-count}(\widetilde{S})/M) \tag{11}$$

where $M = 1$ if Q is absolute, or $M = m = card(\mathcal{D})$ if Q is relative. Assume that n fuzzy sets $\widetilde{S}_1,\ldots,\widetilde{S}_n$ are chosen and at least one of them is of type-2. They represent linguistically expressed properties of objects $y_1, ..., y_m$ described by records $d_1,\ldots,d_m$. The membership function of the type-2 composite summarizer $\widetilde{S} = \widetilde{S}_1$ AND $\widetilde{S}_2$ AND ... AND $\widetilde{S}_n$ is computed as

$$\mu_{\widetilde{S}}(d_i) = \mu_{\widetilde{S}_1 \cap \widetilde{S}_2 \cap \ldots \cap \widetilde{S}_n}(d_i) \tag{12}$$

where the intersection is given by (1). Notice that (12) describes the extension of the George and Srikanth approach [11], and, in consequence, for $n = 1$, also the Yager method of summarization.

3.2 Summaries Based on the Second Canonical Form

Linguistic summaries based on Q^{II}, see (8), are in the form of

$$Q\ P \text{ being } \widetilde{w}_g \text{ are/have } \widetilde{S}\ [T] \tag{13}$$

in which $\widetilde{w}_g$ is represented by a type-2 fuzzy set, and $\widetilde{S}$ is a type-2 or type-1, composite or single summarizer. Similarly to the method presented in [12], the use of the additional fuzzy set enables producing much more interesting summaries. Hence, according to (10), the $\mu_{\widetilde{S}}(d_i)$ is intersected with the membership to the $\widetilde{w}_g$ query:

$$\mu_{\widetilde{w}_g \cap \widetilde{S}}(d_i) = \mu_{\widetilde{w}_g}(d_i) \cap \underbrace{\mu_{\widetilde{S}_1}(d_i) \cap \ldots \cap \mu_{\widetilde{S}_n}(d_i)}_{\mu_{\widetilde{S}}} \tag{14}$$

Step 1. For each $i = 1, ..., m$ compute $\mu_{\widetilde{w}_g}(d_i) \in \mathcal{F}([0,1])$
Step 2. Construct the base $\mathcal{D} \supseteq \mathcal{D}' = \{d_i \colon \mu_{\widetilde{w}_g}(d_i) \neq \emptyset\}$, $m' = card(\mathcal{D}') \leq m$
Hence, the degree of truth of the (13) summary is a real number

$$T = \mu_Q(\text{nf}\sigma\text{-count}(\widetilde{w}_g \cap \widetilde{S})/\text{nf}\sigma\text{-count}(\widetilde{w}_g)) \tag{15}$$

Thanks to Steps 1, 2, the computational cost is reduced from $m \cdot (n+1)$ to at most $m' \cdot n + m$ membership assessments.

3.3 Quality Measures for Type-2 Summaries

This section introduces the original extensions of five measures for type-1 summaries [12]. The next five indices, $T_6 - T_{10}$, are new and specific for type-2 summaries (although their versions for type-1 summaries may also be considered).

1. **Degree of Truth** – see (11), (15).
2. **Degree of Imprecision.** The degree of imprecision of a linguistic summary with a type-2 fuzzy summarizer is determined as

$$T_2 = 1 - \left(\prod_{j=1}^{n} in(\widetilde{S}_j)\right)^{1/n} \tag{16}$$

   The closer to 1 is T_2, the more precise the summary.

3. **Degree of Covering.** The degree of covering is possible to be computed if a summary is based on the second canonical form, see (8).

$$T_3 = card(supp(\widetilde{w}_g \cap \widetilde{S}))/card(supp(\widetilde{w}_g)) \tag{17}$$

   The meaning of the index is the (relative) number of objects corresponding to the query and covered by the summary.
4. **Degree of Appropriateness** – we decompose a summarizer into a number of fuzzy sets $\widetilde{S}_1, \ldots, \widetilde{S}_n$, and for each the r_j index is computed via (11). The degree of appropriateness is based on $g_{i,j}$

$$g_{i,j} = \mu_{supp(\widetilde{S}_j)}(d_i) \tag{18}$$

   which is depends on the support of the $\widetilde{S}_j$ type-2 fuzzy set representing the j-th summarizer. Hence

$$T_4 = \left| \prod_{j=1}^{n} \frac{\sum_{i=1}^{m} g_{i,j}}{m} - T_3 \right| \tag{19}$$

5. **Length of a Type-2 Summary** – depends on $b = card(\{\widetilde{S}_1, \ldots, \widetilde{S}_b\})$ – the number of sets that represent a summarizer, $b \leq n$. The more sets, the less precise the summarizer:

$$T_5 = 2 \cdot (0.5)^b \tag{20}$$

6. **Type-2 Quantification Imprecision** – is analogous to T_2

$$T_6 = 1 - in(Q) \tag{21}$$

7. **Type-2 Quantification Cardinality**

$$T_7 = 1 - card(Q)/N \tag{22}$$

   where $N = 1$ if Q is relative, or $N = card(D(Q))$ if Q is absolute.
8. **Type-2 Summarizer Cardinality**– because of possible several fuzzy sets $\widetilde{S}_1$, ..., $\widetilde{S}_n$ representing the summarizer, the form of T_8 is:

$$T_8 = 1 - \left(\prod_{j=1}^{n} \text{nf}\sigma\text{-count}(\widetilde{S}_j)/card(\mathcal{X}_j) \right)^{1/n} \tag{23}$$

9. **Imprecision of The Type-2 Query** T_9 is determined by the degree of imprecision of the query in a summary based on the second canonical form:

$$T_9 = 1 - in(\widetilde{w}_g) \tag{24}$$

10. **Cardinality of The Type-2 Query**

$$T_{10} = \text{nf}\sigma\text{-count}(\widetilde{w}_g)/card(\mathcal{D}(\widetilde{w}_g)) \tag{25}$$

4 An Improved News Generating

4.1 The Algorithm

The general assumptions of the system which produces compact textual messages from large sets of numerical data, are given in [1]. Below, we present the improved version of the algorithm that generates news with the use of type-2 fuzzy summarizers $\widetilde{S_1}, ..., \widetilde{S_n}$. The measures described in Section 3.3 are applied to select the summaries of the highest goodness (i.e. the most informative).

// generating summaries in the form of Q^I
1. for each non-empty $\hat{S} \subseteq \{\widetilde{S_1}, ..., \widetilde{S_z}\}$
1.1. determine $\mu_{\hat{S}}(d_i)$ via (12)
1.2. for each quantifier Q_h, $h = 1, ..., k$
compute $T_{1,h}$, $T_{6,h}$, and $T_{7,h}$ via (11), (21), and (22), respectively
1.3. compute $T_{h_{\max}} = \max_{h \in \{1,...,k\}} \{t : t = w_1 T_{1,h} + w_6 T_{6,h} + w_7 T_{7,h}\}$, remember $h_{\max}$
1.4. compute T_2, see (16)
// T_3, T_9, T_{10} are not assessed, because of no $\widetilde{w_g}$ queries in Q^I
1.5. compute T_4, via (18), (19), for $T_3 = 0$
1.6. compute T_5 via (20)
1.7. compute T_8 via (23)
1.8. $T = T_{h_{\max}} + w_2 \cdot T_2 + w_4 \cdot T_4 + w_5 \cdot T_5 + w_8 \cdot T_8$
1.9. generate the summary $Q_{h_{\max}}$ P are/have $\hat{S}$ [T]

// generating summaries in the form of Q^{II}
2. for each non-empty query $\widetilde{S_w} \subsetneq \{\widetilde{S_1}, ..., \widetilde{S_z}\}$
and for each non-empty summarizer $\hat{S} \subseteq \{\widetilde{S_1}, ..., \widetilde{S_z}\} \setminus \widetilde{S_w}$
2.1. determine $\mu_{\widetilde{S_w}}(d_i)$ via (14)
2.2. determine $D \supseteq D_w = \{d_i \in D : \mu_{\widetilde{S_w}}(d_i) \neq \emptyset\}$
2.3. for each $d_i \in D_w$ determine $\mu_{\hat{S}}(d_i)$
2.4. for each relative quantifier $Q_h : h \in \{1, ..., k\}$
compute $T_{1,h}$, $T_{6,h}$, and $T_{7,h}$ via (15), (21), and (22), respectively
2.5. compute $T_{h_{\max}}$ analogously to 1.3., remember $h_{\max}$
2.6. compute T_2 analogously to 1.4.
2.7. compute T_3 according to (17)
2.8. compute $T_4 = \left| \prod_{\widetilde{S_j} \in \hat{S}} \frac{\sum_{d_i \in D_w} g_{i,j}}{card(D_w)} - T_3 \right|$, via (18)-(19)
2.9. compute T_5 analogously to 1.6.
2.10. compute T_8 analogously to 1.7.
2.11. compute T_9 and T_{10} via (24), (25), resp.
2.12. $T = T_{h_{\max}} + \sum_{i=2}^{5} w_i \cdot T_i + \sum_{i=8}^{10} w_i \cdot T_i$
2.13. generate summary $Q_{h_{\max}}$ P being S_w are/have $\hat{S}$ [T]

Ad. 1. In this step, finding all non-empty subsets of $\{\widetilde{S_1}, ..., \widetilde{S_z}\}$ is required; the number of such subsets is exactly $2^z - 1$. In the implementation, the problem is resolved via generating binary forms of all natural numbers between 0 and $2^z - 1$. The forms are taken as characteristic vectors of the sought subsets.
Ad. 1.3. and 1.8. $w_1 + w_2 + w_4 + w_5 + w_6 + w_7 + w_8 = 1$.
Ad. 2.10. $w_1 + \ldots + w_{10} = 1$.

4.2 Implementation and Results

The algorithm has been implemented on .NET platform in the C# language. The database (in MS SQL Server (*.mdf) and MS Access (*.mdb) formats, has consisted of ca 10,000 records on workers of a company. The view containing tuples in the form of ⟨`Age, Education, Salary`⟩ has been generated. The summarizers have been determined as values of linguistic variables $L_1 = Age$, $L_2 = Education$, $L_3 = Salary$, e.g. $H(Age) = \{$young, middle-aged, experienced, about 40, about 30$\}$. Each label of L_1, L_3 have been represented by a type-2 fuzzy set, and of L_2 – by crisp sets. Sample results for S_1=about 30, S_2=high school, and S_3=about 4000 is presented:

```
About half of workers are ab. 30 [0.47]. Much more than 2000 workers
graduated from high school [0.74]. About half of workers earn ab.
4000 [0.54]. Many workers graduated from high schools and earn ab.
4000 [0.37]. Many workers graduated from high schools and are ab. 30
[0.38]. Many workers earn ab. 4000 and are ab. 30 [0.37]. Ab. half of
workers graduated from high schools are ab. 30 [0.46].
```

Finally, we notice the results obtained are at least of the same quality that similar given by type-1 summarization methods, see Sec. 5.

5 Evaluating the Success of the Type-2 Summarization

The introduced type-2 linguistic summarization is a generalization of the existing methods based on type-1 fuzzy sets, i.e. summarizers, quantifiers, and queries, are now represented by type-2 membership functions, the values of which are fuzzy numbers. Since a real number is a specific case of fuzzy, type-1 methods can be applied together, because the new approach includes them as specific cases.

However, type-2 membership functions are more complicated than type-1. They are families of at least several type-1 functions that represent given data, e.g. preferences of experts. Unfortunately, they are more time-consuming because more membership values, primary and secondary, must be assessed, see e.g. the definitions of cardinalities for type-1, cf. [10] and type-2 sets, cf. (3).

Hence, although type-2 summaries are more time consuming, we expect that they allow to produce the results that cover also type-1 summaries, in particular, summaries at least as informative as the obtained through type-1 methods, according to the measures of informativeness presented in Sec. 3.3.

Assumptions for comparing type-1 and type-2 summaries. We compare type-1 and type-2 summaries under the following assumptions:

(A1) *The same set of records described by attributes $V_1, \ldots, V_n$ is summarized both under type-1 and type-2 methods.*

(A2) *The $\mathcal{X}_1, \ldots, \mathcal{X}_n$ sets are the domains of $V_1, \ldots, V_n$, respectively, and* $\forall_{i=1,\ldots,n}, \mathcal{X}_i \subseteq R$

(A3) *If a type-1 set A and a type-2 $\widetilde{A}$ in $\mathcal{X}_i$, $i \leq n$, represent the same linguistic term, then A is considered as an embedded type-1 fuzzy set in $\widetilde{A}$*[1].

Comparing type-1 and type-2 quality measures. The quality indices for summaries are based on cardinality and support, (2), (4) for type-1, and (3), (5) for type-2. From these equations, and from the concept of embedded type-1 set:

Proposition 2. *For each type-1 A embedded in type-2 $\widetilde{A}$ in $\mathcal{X}$*

$$card(A) \leq card(\widetilde{A}) \tag{26}$$

Proof. Let $x \in \mathcal{X}$, $u_A = \mu_A(x)$. A is of type-1, hence, $\mu_x(u_A) = 1$. Furthermore, $u_A \in \{u_{\widetilde{A}} \colon \mu_x(u_{\widetilde{A}}) = 1\}$. Thus, $u_A \leq \max\{u_{\widetilde{A}} \colon \mu_x(u_{\widetilde{A}}) = 1\}$, and from (2), (3), we have $\sum_{x\in\mathcal{X}} u_A \leq \sum_{x\in\mathcal{X}} \max\{u_{\widetilde{A}} \in J_x \colon \mu_x(u_{\widetilde{A}}) = 1\}$.

Proposition 3. *For each type-1 A embedded in type-2 $\widetilde{A}$ in $\mathcal{X}$*

$$supp(A) = supp(\widetilde{A}) \;\; \wedge \;\; card(supp(A)) = card(supp(\widetilde{A})) \tag{27}$$

Proof. Let $x \in \mathcal{X}$, $u_A = \mu_A(x), u_A > 0$. Hence $\xi_{supp(A)}(x) = 1$. Since A is of type-1, $\mu_x(u_A) = 1$. Hence, from (5), we have $\mu_{supp(\widetilde{A})} = \sup_{u\in J_x\setminus\{0\}} \mu_x(u) = 1$. Thus, $\forall_{x\in\mathcal{X}}\ \xi_{supp(A)}(x) = \mu_{supp(\widetilde{A})}(x)$.

Proposition 4. *Let (A1)–(A3) are fulfilled. Let type-1 $S_1, ..., S_n, w_g$ in $\mathcal{X}_1, ...,$ $\mathcal{X}_{n+1}$ be embedded in type-2 $\widetilde{S}_1, ..., \widetilde{S}_n, \widetilde{w}_g$ in $\mathcal{X}_1, ..., \mathcal{X}_{n+1}$. Let Q ba a fuzzy quantifier. Let us denote by $T_i(Q, S_1, ..., S_n, w_g)$, and $T_i(Q, \widetilde{S}_1, ..., \widetilde{S}_n, \widetilde{w}_g)$, $i = 1 \ldots 10$, the measures described in Sec 3.3, for Q, S_1,...,S_n, w_g and for Q, $\widetilde{S}_1$,...,$\widetilde{S}_n$, $\widetilde{w}_g$.*

$$T_1(Q, S_1, ..., S_n, w_g) \leq T_1(Q, \widetilde{S}_1, ..., \widetilde{S}_n, \widetilde{w}_g) \;\; \textit{from (26), (11), (15)} \tag{28}$$
$$T_7(Q, S_1, ..., S_n, w_g) \leq T_7(Q, \widetilde{S}_1, ..., \widetilde{S}_n, \widetilde{w}_g) \;\; \textit{from (26), (22)} \tag{29}$$
$$T_8(Q, S_1, ..., S_n, w_g) \leq T_8(Q, \widetilde{S}_1, ..., \widetilde{S}_n, \widetilde{w}_g) \;\; \textit{from (26), (23)} \tag{30}$$
$$T_{10}(Q, S_1, ..., S_n, w_g) \leq T_{10}(Q, \widetilde{S}_1, ..., \widetilde{S}_n, \widetilde{w}_g) \;\; \textit{from (26), (25)} \tag{31}$$

Besides, for $i = 2 \div 6, 9$, $T_i(Q, S_1, ..., S_n, w_g) = T_i(Q, \widetilde{S}_1, ..., \widetilde{S}_n, \widetilde{w}_g)$, see (27).

We conclude from Prop. 4 that the measures based on cardinalities, T_1, T_7, T_8, T_{10} take values greater or equal for type-2 than for type-1 summaries, while measures based on supports, T_2, T_3, T_4, T_6, T_9, take the same values for type-1 and type-2 summaries. Thus, the proposed type-2 summarization allows to achieve the results which are <u>at least as informative</u> as type-1 methods.

---

[1] It represents a proposed type-1 membership function "bridged" with other expert proposals, and, finally, a term is described by a type-2 membership function.

6 Conclusions

The contribution to the domain of data intelligent summarization, presented in this paper, can be, *nomen omen*, s u m m a r i z e d in the following points:

- The original method of linguistic data summarization handled by type-2 fuzzy logic, has been presented.
- The method is an extension of the existing methods based on type-1 fuzzy logic; it covers the previous as a specific case.
- The known quality measures for type-1 summaries have been enhanced to their type-2 versions, and new quality measures of type-2 summaries have been proposed, also applying to type-1 summaries.
- The improved algorithm for finding optimal and the most specific type-2 summaries, has been presented. It is applied to the task and schema presented in [1], and generalizes it.
- The new method produces summaries that are based on more experts preferences. Hence, the results are more informative.

References

1. Niewiadomski, A.: News generating via fuzzy summarization of databases. LNCS, vol. 3831, pp. 419–429. Springer, Heidelberg (2006)
2. Yager, R.: A new approach to the summarization of data. Inf. Sci. 28, 69–86 (1982)
3. Zadeh, L.A.: The concept of linguistic variable and its application for approximate reasoning (i). Information Sciences 8, 199–249 (1975)
4. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Computers and Maths with Applications 9, 149–184 (1983)
5. Mendel, J.M. (ed.): Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions. Prentice-Hall, Upper Saddle River, NJ (2001)
6. Niewiadomski, A.: On two possible roles of type-2 fuzzy sets in linguistic summaries. Lecture Notes in Artificial Intelligence 3528, 341–347 (2005)
7. Niewiadomski, A., Bartyzel, M.: Elements of type-2 semantics in linguistic summaries of databases. Lecture Notes in Artificial Intelligence 4029, 278–287 (2006)
8. Niewiadomski, A., Ochelska, J., Szczepaniak, P.S.: Interval-valued linguistic summaries of databases. Control and Cybernetics (2), 415–444 (2005)
9. Niewiadomski, A., Szczepaniak, P.S.: News generating based on interval type-2 linguistic summaries of databases. In: Proceedings of IPMU 2006 Conference, July 2–7, 2006, Paris, France, pp. 1324–1331 (2006)
10. De Luca, A., Termini, S.: A definition of the non-probabilistic entropy in the setting of fuzzy sets theory. Information and Control 20, 301–312 (1972)
11. George, R., Srikanth, R.: Data summarization using genetic algorithms and fuzzy logic. In: Herrera, F., Verdegay, J. (eds.) Genetic Algorithms and Soft Computing, pp. 599–611. Physica–Verlag, Heidelberg (1996)
12. Kacprzyk, J., Yager, R.R.: Linguistic summaries of data using fuzzy logic. Int. J. of General Systems 30, 133–154 (2001)