

Fuzzy Quantifiers for Data Summarization and their Role in Granular Computing

Ingo Glöckner, Alois Knoll

Department of Technical Computer Science, University of Bielefeld
33501 Bielefeld, P.O.-Box 100131, Germany
{ingo|knoll}@techfak.uni-bielefeld.de

Abstract

Data summarization is an enabling technique of Granular Computing, because of its promise to abstract from individual observations and to view a phenomenon as a whole. The linguistic summaries are built around a fuzzy quantifier which functions as the ‘summarizer’. Linguistic data summarization therefore presupposes an underlying model of fuzzy quantifiers, which is of crucial importance to the adequacy of the generated summaries. In the paper, we present an axiomatic theory of fuzzy quantification. It attempts to formalize the notion of ‘linguistic adequacy’, in order to eliminate the implausible results observed with existing approaches. We provide evidence that the models of the theory are plausible from linguistic considerations. Finally we present three practical models and discuss some of their properties. These models are computational, and systems for data summarization can directly profit from our improvements by plugging in the new algorithms.

1. Introduction

Humans routinely granulate and condense large amounts of data in order to perform everyday tasks and in order to communicate efficiently with others. This is possible because natural language (NL) provides a class of expressions specifically designed to summarise information, viz that of natural language quantifiers. NL quantifiers, and in particular their approximate variety (‘almost all’ etc.), provide flexible means for expressing accumulative properties of collections and can also describe global (e.g., quantitative) aspects of relationships between individuals. The modelling of NL quantifiers might evolve into one of the key techniques for Granular Computing because these provide a view of the phenomenon as a whole, by abstracting from individual observations or data instances. This process is very different from classification and other techniques that

simplify and enhance the description of individual objects, but cannot reduce their sheer number. By contrast, NL quantifiers extract a description of a given collection (e.g. image region, cluster, class) from the descriptions of the individuals gathered in the collection. The original information, scattered over a large number of individuals, is hence distilled into descriptions of only a few collections of interest, based on their accumulative properties. This process of granulating information by utilizing second-order properties, i.e. properties of collections of individuals, or properties of relations between individuals, will be called *summarization*.

2. Data Summarization Based on Fuzzy Quantifiers

Following Zadeh [19], fuzzy set theory attempts to model NL quantifiers by operators called ‘fuzzy quantifiers’. Known models are Zadeh’s Σ -Count approach [20], the FG-Count approach [20, 14], OWA operators [15, 16], and the FE-Count approach [10]. The promise of fuzzy quantifiers for data summarization has soon been recognized. According to Yager [13, 11], there are two relevant types of linguistic summaries, which are either of the type ‘ Q objects in DB are S ’, e.g. ‘Few people in DB are tall’, or of the relative type ‘ Q R objects in DB are S ’, e.g. ‘Most tall objects in DB are young’, where Q is the quantity of agreement (modelled by a fuzzy quantifier); R , S (‘summarizer’) are fuzzy sets representing the NL concepts of interest, and DB is a database.

Today many systems for linguistic data summarization rely on techniques of fuzzy quantification [8, 7, 11, 17]. The basic ideas described above have often been developed into elaborate mechanisms for summary generation and evaluation. However, the underlying models used for interpreting fuzzy quantifiers, have not undergone that rapid evolution; all of the systems listed above still rely on some form of the Σ -count approach

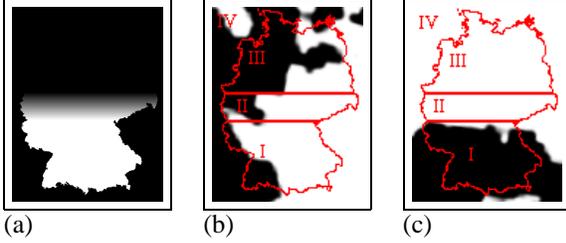


Figure 1. Results of OWA approach for criterion ‘At least 60 percent of Southern Germany are cloudy’. (a) Fuzzy region ‘Southern Germany’, relevant pixels white; (b) desired result: 1, OWA: 0.1; (c) desired result: 0, OWA: 0.6. The results show an undesirable dependency on cloudiness grades in regions III and IV, which do not belong to Southern Germany at all.

or the OWA-approach. This raises severe concerns that the implausible behaviour reported for these models or their extensions, might compromise the quality of the generated summaries [2, 6]. In fact, a rigorous evaluation of existing approaches to fuzzy quantification with respect to their linguistic adequacy has produced negative results in all cases [2], because the traditional models (Σ -count approach, FG-count approach, OWA approach, and FE-count approach) are either too weak (cannot model the interesting cases) and/or subject to gross linguistic failure (i.e. results can be obviously implausible). For an example situation in which one of the approaches fails, consider Fig. 1, which depicts some results obtained from OWA operators with importance qualification [16]. More counter-examples, also covering the other approaches, can be found in [2, 6]. To mention some typical problems, it appears that non-monotonic quantifiers (‘about half of the objects in DB are A ’) are notoriously difficult to most approaches, and extensions to cover such quantifiers usually fail. Most importantly, none of the traditional approaches yields convincing results in the important case of multi-place quantification, e.g. importance qualification like in ‘almost all A in DB are B ’.

The underlying model of fuzzy quantification is of crucial importance to the linguistic adequacy of the generated summaries, but the approaches used for data summarization are faced with severe problems. In the following, we will hence focus on the presentation of an improved theory of fuzzy quantification. This seems to be methodically preferable, and even most promising for practical systems that use fuzzy quantifiers, because it tackles one of core problems of linguistic data summarization. Existing systems can directly profit

from these improvements by simply plugging in the new models.

3. The Linguistic Theory of Fuzzy Quantification

In order to avoid any problems concerning linguistic adequacy, it seems advantageous to profit from the knowledge of linguists, and hence to discuss fuzzy quantifiers in the framework of the *linguistic* theory of quantification. The Theory of Generalized Quantifiers [1] (TGQ) rests on a simple but expressive model of two-valued quantifiers, which provides a uniform representation for absolute and proportional quantifiers, unrestricted and restricted quantification (i.e. involving importances), and even for multi-place quantifiers like ‘more A ’s than B ’s are C ’s’, composite quantifiers like ‘most A ’s and B ’s are C ’s or D ’s’, and non-quantitative examples like ‘John’ or ‘almost all married X ’s are Y ’s’. The theory recognizes more than thirty different types of quantifiers, of which absolute and proportional quantifiers are only two basic examples [9]. Nonetheless, TGQ was not developed with fuzzy sets in mind. Hence all quantifiers and all argument sets involved in the quantification (e.g. weights in importance qualification) must be crisp. TGQ is therefore not (directly) suited for real-world applications like data summarization, which need to deal with imperfect data and non-idealized interpretations of NL concepts, which typically lack clear-cut boundaries.

In the following, we attempt to overcome this limitation by incorporating the notion of fuzziness. We identify a number of cornerstones of a principled theory:

1. Improved Representation through n -ary quantifiers and semi-fuzzy quantifiers

An n -ary fuzzy quantifier \tilde{Q} on a base set $E \neq \emptyset$ assigns to each choice of fuzzy subsets X_1, \dots, X_n of E a gradual result $\tilde{Q}(X_1, \dots, X_n) \in [0, 1]$. Fuzzy quantifiers constitute an expressive class of operators because they introduce a second order construct for fuzzy sets. However, they are often hard to define because the familiar concept of cardinality of crisp sets is not applicable to the fuzzy sets that form the arguments of a fuzzy quantifier. It is necessary to introduce a simplified representation, which must be still powerful enough to embed all quantifiers in the sense of TGQ. An n -ary semi-fuzzy quantifier on a base set E is a mapping which to each choice of crisp subsets Y_1, \dots, Y_n of E assigns a gradual result $Q(Y_1, \dots, Y_n) \in [0, 1]$. Because semi-fuzzy quantifiers must be defined for crisp input only, they

are much easier to define than fuzzy quantifiers. In particular, the usual crisp cardinality is applicable to their arguments and can hence be used to provide an interpretation for semi-fuzzy quantifiers.

2. A *quantifier fuzzification mechanism* (QFM) \mathcal{F} assigns to each semi-fuzzy quantifier Q a fuzzy quantifier $\mathcal{F}(Q)$ of the same arity and on the same base set. These are applicable both to crisp and fuzzy arguments. QFMs are useful because the concepts of TGQ can be easily adapted to the case of semi-fuzzy quantifiers and fuzzy quantifiers. We can then require that a certain property of a quantifier be preserved when applying the QFM, and that \mathcal{F} be compatible with certain constructions on (semi-)fuzzy quantifiers. This can be likened to the well-known mathematical concept of a homomorphism (structure-preserving mapping).
3. We require *compatibility* with concepts of TGQ; an adequate QFM should preserve all properties of linguistic relevance. We enforce this by stating a set of axioms for ‘admissible’ or ‘reasonable’ choices of QFMs, the DFS axioms.
4. We should find *models of the axioms*, i.e. ‘reasonable’ choices of \mathcal{F} , and characterise interesting classes of such models in terms of distinguished properties;
5. *Efficient algorithms* must be developed for implementing the resulting operators.

Following from these requirements, we have developed an axiomatic theory of fuzzy quantification known as ‘DFS theory’. We start the description of our theory by listing the 6 axioms that are required:

1. **Correct Generalisation.** We require that

$$\mathcal{F}(Q)(X_1, \dots, X_n) = Q(X_1, \dots, X_n)$$

whenever X_1, \dots, X_n are crisp (combined with the other axioms, this condition can be restricted to $n \leq 1$).

Rationale: a semi-fuzzy quantifier Q is defined only for crisp arguments, while $\mathcal{F}(Q)$ is defined for arbitrary fuzzy arguments. If all arguments are crisp, Q and $\mathcal{F}(Q)$ must match.

2. **Membership Assessment.** The two-valued quantifier defined by $\pi_e(Y) = 1$ if $e \in Y$ and $\pi_e(Y) = 0$ otherwise for crisp Y , has the obvious fuzzy counterpart $\tilde{\pi}_e(X) = \mu_X(e)$ for fuzzy subsets of E . We require that $\mathcal{F}(\pi_e) = \tilde{\pi}_e$.
Rationale: Membership assessment (crisp or fuzzy) can be modelled through quantifiers. For

an element e of the base set, we can define a two-valued quantifier π_e which checks if e is present in its argument. Similarly, we can define a fuzzy quantifier $\tilde{\pi}_e$ which returns the degree to which e is contained in its argument. It is natural to require that the crisp quantifier π_e be mapped to $\tilde{\pi}_e$, which plays the same role in the fuzzy case.

3. **Dualisation.** We require that \mathcal{F} preserves dualisation of quantifiers, i.e. $\mathcal{F}(Q')(X_1, \dots, X_n) = \tilde{\sim} \mathcal{F}(Q)(X_1, \dots, X_{n-1}, \tilde{\sim} X_n)$ for all fuzzy arguments X_1, \dots, X_n whenever $Q'(Y_1, \dots, Y_n) = \tilde{\sim} Q(Y_1, \dots, Y_{n-1}, \neg Y_n)$ for all crisp arguments Y_1, \dots, Y_n .
Rationale: Obviously, a phrase like ‘all X’s are Y’s’ should have the same result as ‘it is not the case that some X’s are not Y’s’.

4. **Union.** We require that \mathcal{F} preserves unions of arguments, i.e. $\mathcal{F}(Q')(X_1, \dots, X_{n+1}) = \mathcal{F}(Q)(X_1, \dots, X_{n-1}, X_n \cup X_{n+1})$ whenever $Q'(Y_1, \dots, Y_{n+1}) = Q(Y_1, \dots, Y_{n-1}, Y_n \cup Y_{n+1})$.
Rationale: It should not matter whether “many X’s are Y’s or Z’s” is computed by evaluating $\mathcal{F}(\mathbf{many})(X, Y \cup Z)$ or by computing $\mathcal{F}(Q)(X, Y, Z)$ with $Q(X, Y, Z) = \mathbf{many}(X, Y \cup Z)$.

5. **Monotonicity in Arguments.** We require that \mathcal{F} preserve monotonicity in arguments, i.e. if Q is nondecreasing/nonincreasing in the i -th argument, then $\mathcal{F}(Q)$ has the same property. When combined with the other axioms, the condition can be restricted to the case that Q is nonincreasing in its n -th argument).

Rationale: There must be a systematically different interpretation of statements like “all men are tall” and “all young men are tall” where the former statement expresses the stricter condition.

6. **Functional Application.** Finally we require that \mathcal{F} be compatible with a construction called ‘functional application’, i.e. $\mathcal{F}(Q')(X_1, \dots, X_n) = \mathcal{F}(Q)(f'_1(X_1), \dots, f'_n(X_n))$ is valid whenever Q' is defined by $Q'(Y_1, \dots, Y_n) = Q(f_1(Y_1), \dots, f_n(Y_n))$, where f'_1, \dots, f'_n are obtained from the induced extension principle of \mathcal{F} , see [3].
Rationale: This abstract axiom ensures that \mathcal{F} behave consistently over different domains E .

A QFM \mathcal{F} which satisfies these axioms is called a determiner fuzzification scheme, or DFS for short (“determiner” is a synonym from TGQ for “quantifier”). If \mathcal{F} induces the standard negation $\neg x = 1 - x$ and the

standard extension principle of [18], then it is called a *standard DFS*. These DFSes constitute the natural class of standard models of fuzzy quantification.

A large number of properties of linguistic or logical relevance are entailed by the above axioms: If \mathcal{F} is a DFS, then

- \mathcal{F} induces a reasonable set of fuzzy propositional connectives, i.e. $\tilde{\neg}$ is a strong negation $\tilde{\wedge}$ is a t -norm, $\tilde{\vee}$ is an s -norm etc. These connectives are obtained from a canonical construction, see [3];
- $\mathcal{F}(\forall)$ is a T -quantifier and $\mathcal{F}(\exists)$ is an S -quantifier in the sense of [12]. This means that the universal quantifier \forall and the existential quantifier \exists are interpreted plausibly in every DFS.
- \mathcal{F} is compatible with the negation of quantifiers. Hence the equation $\mathcal{F}(Q')(X_1, \dots, X_n) = \tilde{\neg}\mathcal{F}(Q)(X_1, \dots, X_n)$ is valid provided that $Q'(Y_1, \dots, Y_n) = \tilde{\neg}Q(Y_1, \dots, Y_n)$ is valid. For example, the meanings of ‘at least one tall men is lucky’ and ‘it is not the case that no tall man is lucky’ coincide in every DFS;
- \mathcal{F} is compatible with the formation of antonyms. Therefore $\mathcal{F}(Q')(X_1, \dots, X_n) = \mathcal{F}(Q)(X_1, \dots, X_{n-1}, \tilde{\neg}X_n)$ is valid whenever $Q'(Y_1, \dots, Y_n) = Q(Y_1, \dots, Y_{n-1}, \neg Y_n)$. For example, the meanings of ‘every tall men is bald’ and ‘no tall men is not bald’ coincide in every DFS;
- \mathcal{F} is compatible with intersections. This means that equation $\mathcal{F}(Q')(X_1, \dots, X_{n+1}) = \mathcal{F}(Q)(X_1, \dots, X_{n-1}, X_n \tilde{\cap} X_{n+1})$ holds, provided that equation $Q'(Y_1, \dots, Y_{n+1}) = Q(Y_1, \dots, Y_{n-1}, Y_n \cap Y_{n+1})$ holds. For example, the meanings of ‘at least two X ’s are Y ’s’ and ‘the set of X ’s that are Y ’s contains at least two elements’ coincide in every DFS;
- \mathcal{F} is compatible with argument permutations. In other words $\mathcal{F}(Q')(X_1, \dots, X_n) = \mathcal{F}(Q)(X_{\beta(1)}, \dots, X_{\beta(n)})$ is valid whenever $Q'(Y_1, \dots, Y_n) = Q(Y_{\beta(1)}, \dots, Y_{\beta(n)})$, where β is a permutation of $\{1, \dots, n\}$. In particular, symmetry properties of a quantifier are preserved by applying \mathcal{F} . Hence the meaning of ‘about 50 X ’s are Y ’s’ and ‘about 50 Y ’s are X ’s’ coincide in every DFS.
- Finally, \mathcal{F} is compatible with argument insertion. This means that $\mathcal{F}(Q')(X_1, \dots, X_n) = \mathcal{F}(Q)(X_1, \dots, X_n, A)$ is valid whenever $Q'(Y_1, \dots, Y_n) = Q(Y_1, \dots, Y_n, A)$, for a

fixed crisp argument $A \in \mathcal{P}(E)$. For example, the meanings of ‘many (married X)’s are Y ’s’ and ‘(many married) X ’s are Y ’s’ coincide in every DFS.

A number of further important properties are possessed by every DFS. Every DFS is guaranteed to map quantitative (automorphism-invariant) quantifiers like **almost all** or **a few** to quantitative fuzzy quantifiers; and it is guaranteed to map non-quantitative quantifiers like **John** or **most married** to non-quantitative fuzzy quantifiers. In addition, every DFS is *contextual*, which means that the quantification result obtained for $\mathcal{F}(Q)(X_1, \dots, X_n)$ only depends on the behaviour of Q inside the ambiguity ranges $\text{core}(X_i) \subseteq Y_i \subseteq \text{support}(X_i)$, where $\text{core}(X_i)$ denotes the elements with unity membership and $\text{support}(X_i)$ denotes elements with non-zero membership. Every DFS is also known to *preserve extension*, i.e. insensitive to the exact choice of the domain as a whole. For example, we expect that the quantification result of ‘most tall people are bald’ does not depend on the precise choice of the domain, as long as it is large enough to contain the fuzzy subsets **tall** and **bald** of interest. For a comprehensive discussion of adequacy properties of DFSes, see [3].

4. Models for data summarization

We now present three models of the theory, which are suited for data summarization. The model \mathcal{M} (as defined below) is the first DFS that we discovered, and subsequently implemented for application in an experimental retrieval system for multimedia weather documents [5]. The model uses the fuzzy median,

$$\text{med}_{\frac{1}{2}}(u_1, u_2) = \begin{cases} \min(u_1, u_2) & : \min(u_1, u_2) > \frac{1}{2} \\ \max(u_1, u_2) & : \max(u_1, u_2) < \frac{1}{2} \\ \frac{1}{2} & : \text{else} \end{cases}$$

for all $u_1, u_2 \in [0, 1]$. $\text{med}_{\frac{1}{2}}$ can be extended to an operator which accepts arbitrary subsets of $[0, 1]$, viz

$$\text{m}_{\frac{1}{2}} X = \text{med}_{\frac{1}{2}}(\inf X, \sup X),$$

for all subsets of $[0, 1]$. In the following, we need the cut range $\mathcal{T}_\gamma(X) \subseteq \mathcal{P}(E)$ of a fuzzy subset X at the cutting level $\gamma \in [0, 1]$, which corresponds to a symmetrical, three-valued cut of X at γ :

$$\mathcal{T}_\gamma(X) = \{Y \subseteq E : X_\gamma^{\min} \subseteq Y \subseteq X_\gamma^{\max}\}$$

where

$$X_\gamma^{\min} = \begin{cases} X_{\geq \frac{1}{2} + \frac{1}{2}\gamma} & : \gamma \in (0, 1] \\ X_{> \frac{1}{2}} & : \gamma = 0 \end{cases}$$

$$X_\gamma^{\max} = \begin{cases} X_{> \frac{1}{2} - \frac{1}{2}\gamma} & : \gamma \in (0, 1] \\ X_{\geq \frac{1}{2}} & : \gamma = 0 \end{cases}$$

Here $X_{\geq \alpha} = \{e \in E : \mu_X(e) \geq \alpha\}$ denotes α -cut, and $X_{> \alpha} = \{e \in E : \mu_X(e) > \alpha\}$ the strict α -cut. (γ can be thought of as a parameter of ‘cautiousness’.)

By applying the extended fuzzy median to the quantification results obtained for all choices of arguments from the cut ranges, we are now able to interpret fuzzy quantifiers for any fixed choice of the cutting parameter. We hence stipulate

$$Q_\gamma(X_1, \dots, X_n) = m_{\frac{1}{2}}\{Q(Y_1, \dots, Y_n) : Y_i \in \mathcal{T}_\gamma(X_i)\},$$

for all semi-fuzzy quantifiers and fuzzy arguments. In order to define DFSes based on Q_γ , the results obtained at each level of cautiousness must be aggregated. This can be accomplished e.g. by means of integration. We hence define

$$\mathcal{M}(Q)(X_1, \dots, X_n) = \int_0^1 Q_\gamma(X_1, \dots, X_n) d\gamma.$$

It can be shown that \mathcal{M} is a standard DFS. \mathcal{M} is a practical model because it is both continuous in arguments and in quantifiers, i.e. robust against slight changes or noise in the fuzzy arguments X_1, \dots, X_n and in the definition of the quantifier Q .

The integral, which was used in the definition of \mathcal{M} , is not the only possible way of abstracting from the cutting parameter γ . The necessary and sufficient conditions on an aggregation mapping \mathcal{B} which make $\mathcal{M}_{\mathcal{B}}(Q)(X_1, \dots, X_n) = \mathcal{B}((Q_\gamma(X_1, \dots, X_n))_{\gamma \in [0,1]})$ a DFS have been presented in [3]. In the course of this investigation, it turned out that there exists a model \mathcal{M}_{CX} with unique adequacy properties. This model was shown to be the optimal choice from a linguistic perspective, even among the full class of standard models. It is the *only standard model* which permits the compositional interpretation of adjectival restriction by a fuzzy adjective, like in ‘almost all young A’s are B’s’. It is hence guaranteed that $\mathcal{M}_{\text{CX}}(\mathbf{almost\ all\ young})(A, B) = \mathcal{M}_{\text{CX}}(\mathbf{almost\ all})(\mathbf{young} \cap A, B)$. The model also preserves the convexity of absolute quantitative quantifiers like ‘about ten’, ‘between twenty and thirty’. In addition, is continuous in arguments and in quantifiers and hence robust against noise; it propagates fuzziness in arguments and in quantifiers, i.e. less specific input cannot result in more specific outputs. Because of its unique properties, \mathcal{M}_{CX} is the preferred choice for all applications that need to capture NL semantics (additional

properties are discussed in [3]). The model can be defined in terms of the cut ranges and median-based aggregation, but also in the following more compact form.

$$\mathcal{M}_{\text{CX}}(Q)(X_1, \dots, X_n) = \sup\{Q_{V,W}^L(X_1, \dots, X_n) : V_1 \subseteq W_1, \dots, V_n \subseteq W_n\}$$

where

$$Q_{V,W}^L(X_1, \dots, X_n) = \min(\Xi_{V,W}(X_1, \dots, X_n), \inf\{Q(Y_1, \dots, Y_n) : V_i \subseteq Y_i \subseteq W_i\})$$

$$\Xi_{V,W}(X_1, \dots, X_n) = \min_{i=1}^n \min(\inf\{\mu_{X_i}(e) : e \in V_i\}, \inf\{1 - \mu_{X_i}(e) : e \notin W_i\}).$$

Another interesting aspect of \mathcal{M}_{CX} is that it consistently generalises the Sugeno integral and hence the ‘basic’ FG-count approach to arbitrary n -place quantifiers, and to quantifiers that do not fulfill any special monotonicity requirements.

The class of known models has been further broadened by abstracting from the median-based aggregation mechanism that was used to define \mathcal{M} and (implicitly) \mathcal{M}_{CX} . The use of $Q_\gamma(X_1, \dots, X_n)$ in these models is then replaced with a pair of mappings which specify upper and lower bounds of the quantification results obtained for all choices of Y_1, \dots, Y_n in the cut ranges:

$$\begin{aligned} \top_{Q, X_1, \dots, X_n}(\gamma) &= \sup\{Q(Y_1, \dots, Y_n) : Y_i \in \mathcal{T}_\gamma(X_i)\} \\ \perp_{Q, X_1, \dots, X_n}(\gamma) &= \inf\{Q(Y_1, \dots, Y_n) : Y_i \in \mathcal{T}_\gamma(X_i)\}. \end{aligned}$$

In [4], the full class of models definable by $\mathcal{F}_\xi(Q)(X_1, \dots, X_n) = \xi(\top_{Q, X_1, \dots, X_n}, \perp_{Q, X_1, \dots, X_n})$ has been investigated and the necessary and sufficient conditions on ξ have been presented which ensure that \mathcal{F}_ξ be a DFS. Our last example \mathcal{F}_{owa} is representative of this new type of models. It is defined by

$$\mathcal{F}_{\text{owa}}(Q)(X_1, \dots, X_n) = \frac{1}{2} \int_0^1 \top_{Q, X_1, \dots, X_n}(\gamma) d\gamma + \frac{1}{2} \int_0^1 \perp_{Q, X_1, \dots, X_n}(\gamma) d\gamma.$$

\mathcal{F}_{owa} is a standard DFS. The model is of particular interest because it consistently generalises the Choquet integral and hence the ‘basic’ OWA approach, to the ‘hard’ cases of general multiplace and non-monotonic quantifiers. \mathcal{F}_{owa} is a practical model because it is continuous both in arguments and in quantifiers, which ensures a certain stability of the results against noise. However, \mathcal{F}_{owa} does not propagate fuzziness in arguments nor in quantifiers. It is hence inferior to \mathcal{M}_{CX} from an adequacy perspective because less specific input can result in more specific output. Nevertheless, \mathcal{F}_{owa} can be advantageous if the inputs are overly fuzzy and one still needs a fine-grained result ranking, because it can discern cases in which models that propagate fuzziness cease to be informative.

5. Perspective

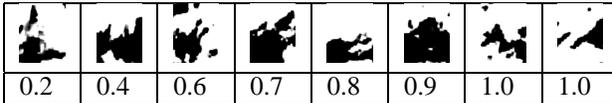
All three models presented above are computational, and algorithms which implement the most common types of quantifiers in these models have been developed [6]. We conclude with an example which demonstrates the suitability of these models to summarize data, see Fig. 2. In this case, the data is comprised of a sequence of images, which depict fuzzy regions of cloudiness grades at given points of time. The images are qualified by importance according to a fuzzy temporal condition, ‘in the last days’. To compute the results, the model \mathcal{M} was used and the quantifier **trp** (for trapezoidal) that was applied is given by

$$\text{trp}_{a,b,c}(Y_1, Y_2) = \begin{cases} t_{a,b}(|Y_1 \cap Y_2|/|Y_1|) & : Y_1 \neq \emptyset \\ c & : Y_1 = \emptyset \end{cases}$$

$$t_{a,b}(z) = \begin{cases} 0 & : z < a \\ \frac{z-a}{b-a} & : a \leq z \leq b \\ 1 & : z > b. \end{cases}$$

The use of the improved models presented here is also encouraged for future systems that generate linguistic summaries from relational data.

Data:



Results:

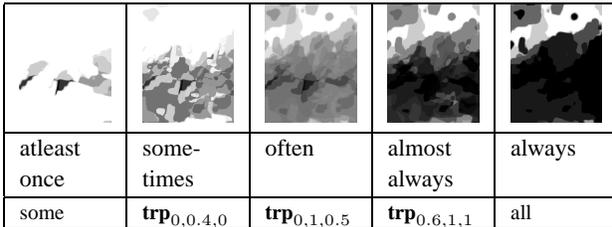


Figure 2. Image sequence and summarization results for various choices of the criterion Q-times cloudy in the last days. Regions that meet the criterion are depicted white.

References

- [1] J. Barwise and R. Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.
- [2] I. Glöckner. A framework for evaluating approaches to fuzzy quantification. TR99-03, Technische Fakultät, Universität Bielefeld, P.O.-Box 100131, 33501 Bielefeld, Germany, 1999.
- [3] I. Glöckner. Advances in DFS theory. TR2000-01, Technische Fakultät, Universität Bielefeld, P.O.-Box 100131, 33501 Bielefeld, Germany, 2000.
- [4] I. Glöckner. A broad class of standard DFSes. TR2000-02, Technische Fakultät, Universität Bielefeld, P.O.-Box 100131, 33501 Bielefeld, Germany, 2000.
- [5] I. Glöckner and A. Knoll. Natural-language navigation in multimedia archives: An integrated approach. In *Proceedings of the Seventh ACM Multimedia Conference (MM '99)*, pages 313–322, Orlando, Florida, 1999.
- [6] I. Glöckner and A. Knoll. A formal theory of fuzzy natural language quantification and its role in granular computing. In W. Pedrycz, editor, *Granular Computing: An Emerging Paradigm*. Physica-Verlag, 2001.
- [7] J. Kacprzyk. Intelligent data analysis via linguistic data summaries: A fuzzy logic approach. In R. Decker and W. Gaul, editors, *Classification and Information Processing at the Turn of the Millenium*, pages 153–161. Springer, 2000.
- [8] J. Kacprzyk and P. Strykowski. Linguistic summaries of sales data at a computer retailer: A case study. In *Proc. of 8th Int. Fuzzy Systems Assoc. World Congress (IFSA '99)*, pages 29–33, 1999.
- [9] E. Keenan and J. Stavi. A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9, 1986.
- [10] A. Ralescu. A note on rule representation in expert systems. *Information Sciences*, 38:193–203, 1986.
- [11] D. Rasmussen and R. Yager. A fuzzy SQL summary language for data discovery. In D. Dubois, H. Prade, and R. Yager, editors, *Fuzzy Information Engineering*, pages 253–264. Wiley, 1997.
- [12] H. Thiele. On T-quantifiers and S-quantifiers. In *The Twenty-Fourth International Symposium on Multiple-Valued Logic*, pages 264–269, Boston, MA, 1994.
- [13] R. Yager. A new approach to the summarization of data. *Information Sciences*, 28:69–86, 1982.
- [14] R. Yager. Approximate reasoning as a basis for rule-based expert systems. *IEEE Trans. on Syst., Man, and Cybernetics*, 14(4):636–643, 1984.
- [15] R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst., Man, and Cybernetics*, 18(1):183–190, 1988.
- [16] R. Yager. Connectives and quantifiers in fuzzy sets. *Fuzzy Sets and Systems*, 40:39–75, 1991.
- [17] R. Yager and J. Kacprzyk. Linguistic data summaries: A perspective. In *Proc. of 8th Int. Fuzzy Systems Assoc. World Congress (IFSA '99)*, pages 44–48, 1999.
- [18] L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Sciences*, 8,9:199–249,301–357, 1975.
- [19] L. Zadeh. A theory of approximate reasoning. In J. Hayes, D. Michie, and L. Mikulich, editors, *Machine Intelligence*, volume 9, pages 149–194. Halstead, New York, 1979.
- [20] L. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computers and Mathematics with Applications*, 9:149–184, 1983.