New Approach to Decision Making

Magdalena Kacprzak¹, Bartłomiej Starosta², and Katarzyna Węgrzyn-Wolska³

¹ Bialystok University of Technology, Bialystok, Poland
² Polish-Japanese Academy of Information Technology, Warsaw, Poland
³ Allianastic, ESIGETEL, Villejuif, France
m.kacprzak@pb.edu.pl, barstar@pjwstk.edu.pl,
katarzyna.wegrzyn@esigetel.fr

Abstract. The aim of this work is to propose a new methodology for representation and processing of imprecise ideas like *good hotel, good food*, etc. As an example we consider the problem of verification of a hotel's features, to choose the best place for a conference. In this approach, a vague term of *a good hotel for a conference* is modeled in paradigm of metasets, which involves splitting it into a tree-like hierarchy of related sub-concepts. Then, the techniques of Opinion Mining are used to determine the polarity and the intensity of opinions available on the Internet, that concern specific hotels. Finally, the membership grades for the sample hotels are calculated. These grades reflect each hotel's total capabilities and represent the degree to which the vague idea is satisfied by each. The result is a new system that helps in decision making.

Keywords: metaset, partial membership, opinion mining

1 Introduction

In many applications offering solutions to real-life problems there is a need for dealing with vague and imprecise data. Such terms often appear in natural language but are hard to represent and process in software tools and formal, mathematical models. Various approaches to this problem exist in the theory of fuzzy sets [15] or rough sets [9]. Similar, but still different approach is proposed by the theory of metasets [10]. The general idea of metaset is inspired by the method of forcing [1] in the classical set theory [5,8]. Metasets admit partial membership, partial equality and other set-theoretic relations [10] which may be evaluated in a Boolean algebra. The certainty values for metaset relations or even compound sentences [13,12] may also be represented as natural language terms.

The main objective of our work is to build a decision support system that will define linear order in the set of objects taking into account characteristic which is vague and unclear. Such an idea can be for example: good restaurant, interesting place to visit, etc. More specifically, having a set of certain objects, e.g. restaurants or bars, there is a need to collate them linearly from best to worst. It is a difficult and non-trivial task because: (a) criteria for evaluation may be varied, (b) it is difficult to compare some characteristics (e.g. friendly service and good food), (c) criteria often depend on the subjective preferences of customers (some prefer crowded rooms with loud music others prefer peaceful and romantic places). In this work, as an example, we consider the search for a suitable hotel for conferences and we will therefore focus on imprecise expression: a good hotel for a conference. On many websites there are different descriptions of hotels available in a selected village. They are often incomplete, inaccurate or simply confusing. Hotel search engines like: www.booking.com or www.trivago.com are coming to our aid by providing opinions of people who visited these hotels. In this article we introduce a basis for decision support system that will help to evaluate hotels. We achieve this goal by extracting the most important features and then representing them as metasets. Having representation of the term a good hotel for a conference in the form of metaset, we can evaluate specific objects. First, opinions about hotels are gathered. We use methods and techniques of Opinion Mining and on their basis we make hotel verification. The resulting degree is a numeric value which helps to verify and compare different hotels.

One of the most important tasks of our research is to analyze various imprecise concepts with the help of metasets and then to explore potential applications. The use of the theory of metasets for modeling vague terms is described in detail in [6]. In this approach, representing any imprecise concept with a metaset lies in splitting it into a tree-like hierarchy of related sub-concepts. In [6] we show how the imprecise idea of a *perfect holiday destination* is represented as a metaset of places whose membership degrees in the metaset are interpreted as their qualities. Then we use it to model and solve the problem of evaluation of the attractiveness of tourist destinations, which is helpful in solving Tourist Trip Design Problems, TTDP.

Another important aspect is the problem of modeling human attitudes towards imprecise ideas. We deal with this problem in [7], where our mechanism is applied to solve the problem of selecting the car best matching the imprecise idea of *a good car for a lady*. In this case, Opinion Mining techniques are applied to build a preference function which reflects someone's attitude towards some imprecise idea. In this approach, the outcome of the decision support tool is a numerical score assigned to a specific car, which expresses how much this car fits the needs of the client. The obtained results confirm his/her opinions shared on the Internet.

In the current paper, the focus is on verification of hotel capabilities in confrontation with the opinions available on the Internet. That's why we do not deal with the subjective preferences of one person, and we investigate whether the information placed on the official website of the hotel are confirmed in opinions of persons who have benefited from its services. For automatic analysis of posts placed on the Web, Opinion Mining techniques are applied. They help to determine polarity and intensity of a given text, i.e., whether it is positive, negative, or neutral and to what extent. To classify the intensity of opinions, we use methods introduced in [2,3,4]. As a result, we propose a new methodology that can be used to develop a mechanism for assigning specific numerical values to objects and on this basis to classify them.

2 Metasets

We need a tool for modeling partial satisfaction of the property *a good hotel for a conference* by different hotels. We use for this purpose a metaset – a set with partial membership relation [10,11]. Metasets are similar to fuzzy sets [15] or rough sets [9]. A metaset may represent an imprecise notion just like a classical set represents a sharp property: it consist of elements which satisfy the property. In the case of a metaset the property may be satisfied to a degree other than full certainty. The degrees are represented by nodes of the binary tree and they may be evaluated as real numbers.

2.1 Basic Definitions

By the definition, a metaset is a relation between elements of some given set and the nodes of the binary tree \mathbb{T} . For simplicity, in this paper we deal with finite first-order metasets only, which we call metasets anyway.

Definition 1. A set which is either the empty set \emptyset or which has the form:

$$\tau = \{ \langle \sigma, p \rangle : \sigma \text{ is a set, } p \in \mathbb{T} \}$$

is called a first-order metaset.

Informally, we interpret the pair-like structure so that the second elements of pairs are nodes of \mathbb{T} which determine membership degrees of members which are first elements of pairs.

The binary tree \mathbb{T} is the set of all finite binary sequences. It is ordered by the reverse prefix relation: if $p, q \in \mathbb{T}$ and p is a prefix of q, then $q \leq p$ (see Fig. 1). The root 1, which is the largest element of \mathbb{T} in this ordering, is the empty sequence.

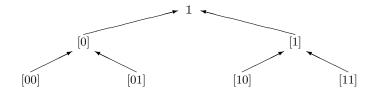


Fig. 1. The levels $\mathbb{T}_0 - \mathbb{T}_2$ of the binary tree \mathbb{T} and the ordering of nodes. Arrows point at the larger element.

The binary sequences which are members of \mathbb{T} are denoted with square brackets, e.g.: [00], [101]. If $p \in \mathbb{T}$, then we denote its direct descendants with $p \cdot 0$ and $p \cdot 1$. A *level* \mathbb{T}_n in \mathbb{T} is the set of all finite binary sequences with the same length n. The level 0 consists of the empty sequence 1 only. A *branch* in \mathbb{T} is an infinite binary sequence. We write $p \in C$ meaning that the binary sequence $p \in \mathbb{T}$ is a prefix of the branch C.

2.2 Interpretations

A metaset determines a collection of classical crisp sets which are called interpretations of the metaset. Each interpretation represents some particular, sharp point of view on the source metaset. For instance, there may be many particular, concrete approaches to the vague idea of *a good hotel for a conference*, shared by different experts. Interpretations of metasets are produced using branches of the binary tree.

Definition 2. Let τ be a first-order metaset and let C be a branch. The set

 $\tau_{\mathcal{C}} = \{ \sigma \in \operatorname{dom}(\tau) \colon \langle \sigma, p \rangle \in \tau \land p \in \mathcal{C} \}$

is called the interpretation of the first-order metaset τ given by the branch C.

In the above definition dom $(\tau) = \{ \sigma : \exists_{p \in \mathbb{T}} \langle \sigma, p \rangle \in \tau \}$ is the domain of τ .

Informally, interpreting a metaset with the given branch C involves two steps. The first step removes all the ordered pairs whose second elements are nodes which do not belong to the branch C. The second stage replaces the remaining pairs – whose second elements lie on the branch C – with their first elements.

2.3 Partial Membership

A member may belong to a metaset to a variety of degrees which are represented by nodes of \mathbb{T} and which in fact make up a Boolean algebra. This property extends the classical case of two-valued membership relation and therefore enables modeling of imprecision. Membership and other set-theoretic relations for metasets are defined using interpretations.⁴

Definition 3. We say that the metaset σ belongs to the metaset τ under the condition $p \in \mathbb{T}$, whenever for each branch C containing p holds $\sigma_{\mathcal{C}} \in \tau_{\mathcal{C}}$. We use the notation $\sigma \epsilon_p \tau$.

Formally, we define infinitely many membership relations: each $p \in \mathbb{T}$ specifies another relation ϵ_p . Any two metasets may be simultaneously in multiple membership relations qualified by different nodes: $\sigma \epsilon_p \tau \wedge \sigma \epsilon_q \tau$. The relation ϵ_1 is equivalent to full, unconditional membership of crisp sets.

The conditional membership reflects the idea that an element σ belongs to a metaset τ whenever some conditions are fulfilled. The conditions are described by the nodes of \mathbb{T} .

Example 1. The ordinal number 1 is the set $\{0\}$ and 0 is just the empty set \emptyset . Let $\tau = \{\langle 0, [0] \rangle, \langle 1, [1] \rangle\}$ and let $\sigma = \{\langle 0, [1] \rangle\}$. Let $\mathcal{C}^0 \ni [0]$ and $\mathcal{C}^1 \ni [1]$ be arbitrary branches containing [0] and [1], respectively. Interpretations are: $\tau_{\mathcal{C}^0} = \{0\}, \tau_{\mathcal{C}^1} = \{1\}, \sigma_{\mathcal{C}^0} = 0$ and $\sigma_{\mathcal{C}^1} = \{0\} = 1$. We see that $\sigma \epsilon_{[0]} \tau$ and $\sigma \epsilon_{[1]} \tau$. Also, $\sigma \epsilon_1 \tau$ holds.

Note, that even though interpretations of τ and σ vary depending on the branch, the metaset membership relation is preserved.

4

⁴ For the detailed discussion of the relations or their evaluation the reader is referred to [11] or [13].

2.4 Evaluating Membership

In applications we usually need numerical evaluation of membership relation. We define it in two steps. First, we consider the smallest subset of \mathbb{T} consisting of elements which determine the membership.

Definition 4. Let σ , τ be first-order metasets. The set

 $\|\sigma \in \tau\| = \max\{p \in \mathbb{T} : \sigma \epsilon_p \tau\}$

is called the certainty grade for membership of σ in τ .

Here, max { $p \in \mathbb{T}$: $\sigma \epsilon_p \tau$ } denotes the set of maximum elements (in the tree ordering) of the set of nodes in \mathbb{T} , for which the relation $\sigma \epsilon_p \tau$ holds.

Now we define the numerical evaluation of membership taking the following assumptions. All the nodes within a level contribute equally to the membership value – none of them is distinguished. For the given $p \in \mathbb{T}$, its direct descendants $p \cdot 0$ and $p \cdot 1$ add half of the contribution of the parent p, each. Therefore, the contribution of a $p \in \mathbb{T}$ must be equal to $\frac{1}{2^{|p|}}$, where |p| is the length of the sequence p.

Definition 5. Let σ , τ be first-order metasets. The following value is called the certainty value of membership of σ in τ :

$$|\sigma \in \tau| = \sum_{p \in \|\sigma \in \tau\|} \frac{1}{2^{|p|}}$$

One may easily see that $|\sigma \in \tau| \in [0, 1]$. If $||\sigma \in \tau|| = \{1\}$, i.e., $\sigma \in \tau_1$ holds, then $|\sigma \in \tau| = 1$. And if $||\sigma \in \tau|| = \emptyset$ ($\sigma \in \rho$ holds for no p), then $|\sigma \in \tau| = 0$.

2.5 Representing Imprecise Ideas with Metasets

Just like a set represents a collection of objects which satisfy a property given by a formula, a metaset represents a "fuzzy" collection of objects which satisfy some imprecise idea. In this paper we use a metaset to represent the imprecise term of *a good hotel for a conference*. Its members are particular hotels which match the given idea to a variety of degrees, usually different than the complete truth.

The core of the idea of representing any imprecise concept with a metaset lies in splitting it into a tree-like hierarchy of related sub-concepts. In this particular case, we require that a good conference hotel has proper facilities for users (user) and organizers (org). From a user point of view these facilities must assure a place for relaxation (relax) and professional work (work). Taking organizer's perspective we must assure proper rooms (room) and equipment. Participants want to relax in a bar or in spa and fitness locations (spa). For professional work during the conference they need good network access (net) and writing-desks in rooms for preparing notes and papers (desk). The organizers must provide large halls for plenary lectures (*large*) and several medium-sized places for parallel sessions (*medium*). The required *equipment* includes projectors and screens, audio sound systems, etc. (*furniture*) as well as minor office accessories like printers, photocopiers, faxes, etc. (*accessory*).

We may continue the process of splitting notions further to achieve arbitrary high precision of description, however, we stop at the 3rd level, for the sake of simplicity.

Note, that by the above description, a parent node stands for the union of two children nodes. For instance, if a hotel assures *relax*, then it means that both *net* and *desk* property are satisfied. And when *furniture* and *accessory* are present, then we say understand that *equipment* is there.

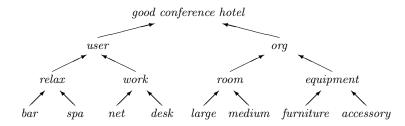


Fig. 2. The binary tree of the features describing a good hotel for a conference

The binary tree in Fig. 2 is used throughout the paper to represent the discussed idea of a good hotel for a conference by means of the metaset Δ . Note, that the nodes of the tree which determine the membership degrees are natural language terms, which also describe reasons for some particular hotel to satisfy the discussed idea.

3 Validation by Opinion Mining Techniques

More and more people share their opinions on various topics on the Internet. We can find in this way the views and experiences of a very large number of people who are not our friends, nor experts in the field, but people who may have the same tastes as ours, so their opinions can be very helpful before making our choice and to have our own ideas on a given topic. It may also consist of the evaluations of specific aspects of the product (capabilities of *a good hotel for a conference*), and feedback from readers about the official data.

Our objective is to create a system to evaluate and validate the typical choice of hotel's reservation for the conference. The main challenge is to extract from the consumers' reviews the most important opinions concerning selected hotel's features and to evaluate these opinions to validate the official data about the hotel. To perform this task we use Opinion Mining techniques.

3.1 Opinion Mining: Basic Definition

Opinion Mining consists in identifying orientation or intensity of opinion. It enables determining whether some part of document or a sentence expresses positive, negative or neutral opinion towards some object (O) or more. Also, it allows for classification of opinions according to intensity degrees.

Definition 6. An opinion is a quadruple (O, F, H, S), where O is a target object, $F = \{f_1, f_2, \ldots, f_n\}$ is a set of features of the object O, H is a set of opinion's holders, S is the set of sentiment/opinion values of the opinion's holder on the feature f_i of the object O.

An object O is represented with a finite set of features, $F = \{f_1, f_2, \ldots, f_n\}$. Each feature $f_i \in F$ $(i = 1, \ldots, n)$ can be expressed with a finite set of words or phrases W_i , where each $W_i = \{W_{i1}, W_{i2}, \ldots, W_{in}\}$ is a set of corresponding synonyms for the features.

Thus, an object O is represented as a tree or taxonomy of components F (or parts), sub-components, and so on. Each node represents a component and is associated with a set of attributes. O is the root node, which also has a set of attributes. An opinion can be expressed on any node or attribute of the node.

In general, the first step of such a process is to retrieve the information from the Web [14] (specialized websites, tweets, blogs, forums, etc.) related to the object ($O: a \ good \ hotel \ for \ a \ conference$ in our case, presented in Ex. 2), to extract the opinions about the selected features (F: nodes of the binary tree) and then to classify this information according to their emotional value (polarity).

The classification of the opinion polarity consists in making decision between positive and negative status. A value called semantic orientation is created in order to demonstrate words' polarity. There are several calculation methods of the words semantic orientation (SO). The most often used method is called SO-A (Semantic Orientation from Association):

$$SO-A(word) = \Sigma_{p \in P} A(word, p) - \Sigma_{n \in N} A(word, n)$$

where:

- -A(word, p) is the association of studied word with the positive word,
- -A(word, n) is equivalent negative,
- -A(word) is a measure of association.

If the sum is positive, the word is oriented positively, and if the sum is negative, the orientation is negative. The absolute value of the sum indicates the orientation intensity.

3.2 Validation Methodology

The purpose is to find the opinions related to the selected features, which could confirm or reject the candidate for *a good hotel for a conference*.

The general idea is the following: find on the specialized websites (booking.com, hotels.com and trivago.com) the list of potential candidates O corresponding to our needs (important features), retrieve for each hotel from candidates' list the set A of the capabilities declared in official hotel's description $(A \subset F)$, extract the clients' opinions corresponding to selected features $f_i \in F$ (i = 1, ..., n) and then validate the hotel's selection.

In this way the positive/neutral opinion on existing feature f_i which belongs to A validates f_i , the positive opinion on feature f_i which does not belong to A adds new feature f_i to A, the negative one cancels f_i from A set. Therefore, the polarity of opinion (negative, positive) is considered to be the significant contribution which evaluates the selection.

In order to demonstrate our methodology we use the hierarchy of conditions depicted in Fig. 2, which comprise the notion of a good hotel for a conference. By Def. 6, the basic components of an opinion are: object O (on which an opinion is expressed, in our case it is a hotel), feature f_i describing the object O, and sentiment/opinion (a view, attitude, or appraisal on each selected feature). According to the idea presented by the tree in Fig. 2, the set F is composed of 14 elements {user, relax, bar, spa, work, net, desk, org, room, large, medium, equipment, furniture, accessory}. For each element $f_i \in F$ we select manually the corresponding set W_i . For example:

To classify intensity of opinions concerning hotel guests, we use the engine of our system [2,3,4]. We formalize our methodology in the following example.

Example 2. For demonstration purposes we consider here opinions on one selected hotel H. We know from the official website that its capabilities include: big conference room (large), desks in rooms (desk), Internet access (net), fitness locations (spa), bar (bar). So at this stage we assume for this hotel the following capabilities: large, bar, spa, desk, net.

To confirm these data we analyze the guest reviews about it retrieved from booking.com and trivago.com websites. The posts include opinions of this sort: "seats in conference room were not comfortable" (*-furniture*), "not able to regulate air condition" (*-furniture*), "very easy to connect my iPad to video projector" (*+accessory*), "speed color printer" (*+accessory*), "I had Wi-Fi access only near the reception" (*-net*), "fitness has always been busy" (*-spa*), "poor sport equipment" (*-spa*), "nice bar" (*+bar*).

Based on these we have validated/canceled the following capabilities for the hotel H: large, accessory, bar, desk.

4 Modeling with Metasets

We model the property of a good hotel for a conference with a metaset and its partial membership relation: if a hotel σ satisfies a property $p \in \mathbb{T}$ (see Fig. 2), then we formally write $\sigma \epsilon_p \Delta$.

By utilizing web retrieval information and opinion mining techniques we discovered that the sample hotel α has a bar, good network access, large conference rooms and the appropriate furniture, and another hotel β has proper places for relaxation and professional work, all types of rooms necessary, and is well equipped with minor office accessories:

$$\alpha: bar, net, large, furniture, \qquad (1)$$

$$\beta$$
: relax, work, room, accessory. (2)

Formally we write:

$$\alpha \epsilon_{bar} \Delta \wedge \alpha \epsilon_{net} \Delta \wedge \alpha \epsilon_{large} \Delta \wedge \alpha \epsilon_{furniture} \Delta, \qquad (3)$$

$$\beta \epsilon_{relax} \Delta \wedge \beta \epsilon_{work} \Delta \wedge \beta \epsilon_{room} \Delta \wedge \beta \epsilon_{accessory} \Delta.$$
(4)

By the Def. 3 we also conclude that:

$$\beta \epsilon_{user} \Delta \wedge \beta \epsilon_{room} \Delta \wedge \beta \epsilon_{accessory} \Delta.$$
(5)

We calculate membership grades for the sample hotels (cf. Def. 4). The grades reflect each hotel's total capabilities and represent the degree to which the vague idea is satisfied by each.

$$\|\alpha \in \Delta\| = \{ bar, net, large, furniture \} , \tag{6}$$

$$= \{ [000], [010], [100], [110] \} , \tag{7}$$

$$\|\beta \in \Delta\| = \{ user, room, accessory \} , \tag{8}$$

$$= \{ [0], [10], [111] \} .$$
(9)

The numerical evaluation of the hotels' capabilities represented by conditional membership in Δ is carried out as follows (cf. Def. 5):

$$|\alpha \in \Delta| = \frac{1}{2^{|[000]|}} + \frac{1}{2^{|[010]|}} + \frac{1}{2^{|[100]|}} + \frac{1}{2^{|[110]|}} , \qquad (10)$$

$$= \frac{1}{2^3} + \frac{1}{2^3} + \frac{1}{2^3} + \frac{1}{2^3} = \frac{4}{8} = 0.5, \qquad (11)$$

$$|\beta \in \Delta| = \frac{1}{2^{|[0]|}} + \frac{1}{2^{|[10]|}} + \frac{1}{2^{|[111]|}}, \qquad (12)$$

$$= \frac{1}{2^2} + \frac{1}{2^1} + \frac{1}{2^3} = \frac{7}{8} = 0.875.$$
 (13)

The results indicate, that β is a member of Δ to a higher degree than α . In other words, β satisfies the property *a good hotel for a conference* better than α . Therefore, the former is the better choice and it is recommended by the discussed system.

5 Conclusions

This paper shows for the first time how the paradigm of metasets and Opinion Mining techniques can be combined to obtain an excellent methodology to validate the information posted on the Internet. This is especially troublesome when the information relates to vague terms and concepts. Our approach can be applied in software decision support systems that help users to make decisions. In this paper the decision involves choosing a good hotel for a conference but can be easily applied to other similar problems.

Acknowledgments Kacprzak's contribution was supported by the Bialystok University of Technology grant S/W/1/2014.

References

- 1. Cohen, P.: The Independence of the Continuum Hypothesis 1. Proceedings of the National Academy of Sciences of the United States of America 50, 1143–1148 (1963)
- Dziczkowski, G., Wegrzyn-Wolska, K.: Rcss rating critics support system purpose built for movies recommendation. In: Advances in Intelligent Web Mastering. Springer (2007)
- Dziczkowski, G., Wegrzyn-Wolska, K.: An autonomous system designed for automatic detection and rating of film. Extraction and linguistic analysis of sentiments. In: Proceedings of WIC, Sydney (2008)
- 4. Dziczkowski, G., Wegrzyn-Wolska, K.: Tool of the intelligence economic : Recognition function of reviews critics. In: ICSOFT 2008 Proceedings.
- 5. Jech, T.: Set Theory: The Third Millennium Edition, Revised and Expanded. Springer-Verlag, Berlin Heidelberg New York (2006)
- Kacprzak, M., Starosta, B.: An Approach to Making Decisions with Metasets. In Trends in Contemporary Computer Science, Podlasie 2014, 159–172 (2014)
- Kacprzak, M., Starosta, B., Wegrzyn-Wolska, K.: Metasets and opinion mining in new decision support system. LNCS vol. 9120(2), pp. 625-636 (2015).
- 8. Kunen, K.: Set Theory, An Introduction to Independence Proofs. Studies in Logic and Foundations of Mathematics, North-Holland Publishing Company (1980)
- Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences 11, 341–356 (1982).
- Starosta, B.: Metasets: A New Approach to Partial Membership, LNAI 7267 pp. 325–333. Springer-Verlag (2012)
- Starosta, B., Kosiński.W: Meta Sets. Another Approach to Fuzziness. In: Seising, R. (ed.) Views on Fuzzy Sets and Systems from Different Perspectives. Philosophy and Logic, Criticisms and Applications. STUDFUZZ, vol. 243, pp. 509–522 (2009)
- Starosta, B., Kosiński, W.: Metasets, Intuitionistic Fuzzy Sets and Uncertainty. LNAI 7894 pp. 388–399. Springer-Verlag (2013)
- Starosta, B., Kosiński, W.: Metasets, Certainty and Uncertainty. In: Atanassov, K.T. et al. (eds.), New Trends in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics. Volume I: Foundations, pp. 139–165. SRI PAS (2013)
- Wegrzyn-Wolska, K., Bougueroua, L.: Tweets mining for French Presidential Election, In Proc. of CASoN 2012, SaO Carlos, Brazil, November (2012)
- 15. Zadeh, L. A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)

10