Metasets and Opinion Mining in New Decision Support System

Magdalena Kacprzak[1], Bartłomiej Starosta[2], and Katarzyna Węgrzyn-Wolska[3]

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

Abstract. The paper is devoted to the problem of modeling human attitudes towards imprecise ideas. A metaset is used for representing an imprecise concept and Opinion Mining techniques are applied to build a preference function which reflects someone's attitude towards the idea. The preferences are then evaluated as real numbers for the sake of comparison and selection of the best matching instance. 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. The nodes of the tree determine the membership degrees for metaset members and they are natural language terms which also describe reasons for some particular member to satisfy the represented idea. The Opinion Mining allows for automatic gathering and evaluation of opinions from the Internet. The proposed mechanism is applied to solve the problem of selecting the car best matching the imprecise idea of *a good car for a lady*. This approach can be applied in a decision support systems that helps both marketers and customers.

Keywords: metaset, partial membership, opinion mining

1 Introduction

In the contemporary marketing two issues play an important role: to identify customer preferences and desires, and to select the product that fits the customer's needs. On the one hand, sellers and manufacturers want to know what might appeal to a potential client, who are potential customers for their products and how to select the product which meets customers' expectations. On the other hand, in the age of rapidly evolving technologies, a producer is the one who awakens in the client the need to have a new model of a smartphone or other kind of mobile devices. Product advertising needs to hit the preferences of users to be effective. This is why it is very important to recognize the opinions of various people about the product. The aim of our research is to design a system that can help in making decisions for both the customer and the manufacturer. More specifically, we are working on a system that will: (a) collect people's opinions about the product and select the features of the product which are the most important for them; (b) create a profile of a customer or a group of customers; (c) calculate (evaluate) how much the selected product matches the customer profile. In our approach we do not focus on commonly shared opinions only. We rather put emphasis on the individual needs of users, especially if they are unusual. The aim of the analysis is to determine how much the chosen product fits the selected user. The obtained result is described with numerical values.

In the paper [7] we showed how to use the idea of metasets to model and solve the problem of evaluation of the attractiveness of tourist destinations. In this case, 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. Client preferences are functions which enable real-number evaluation of the subjective rating of a given destination. The input in this problem is a list of sites with the location and a brief description of each. The output has to be a numeric score assigned to each location that allows us to compare them and ultimately select the best one. Such an approach can be used in automated personalized tour-planning devices. In particular, it can be used in solving Tourist Trip Design Problems, TTDP (see e.g. [15]).

Metasets are the perfect tool for representation and processing of vague, imprecise data, similarly to fuzzy sets [17] or rough sets [9]. Metasets admit partial membership, partial equality and other set-theoretic relations [11] which may be evaluated in a Boolean algebra. The certainty values for metaset relations or even compound sentences [14,13] may also be represented as natural language terms, what is especially important in applications. The general idea of metaset is inspired by the method of forcing [1] in the classical set theory [8,6]. Despite these abstract origins, the definitions of metaset and related notions (i.e. set-theoretic relations or algebraic operations) are directed towards efficient computer implementations [12] and applications [10,7].

In the current paper we develop another application of metaset concept. We use a metaset for representing the imprecise term of *a good car for a lady*. For this metaset we acquire data which are used for building the preference function for a sample client. This function is a slight modification of membership evaluation function for metasets. The data are acquired using the methods and techniques of Opinion Mining. They involve building a system to collect and categorize opinions about a product. This consists in examining natural language conversations happening around a certain product for tracking the mood of the public. The analysis is performed on large collections of texts, including web pages, on-line news, Internet discussion groups, on-line reviews, web blogs, and social media. Opinion Mining aims 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].

The paper is structured as follows. In Sec. 2 we briefly recall the main definitions and lemmas concerning metasets. Section 3 is devoted to issues of Opinion Mining. Section 4 presents the problem of detection of users' preferences. Section 5 gives the solution to the problem in terms of metasets. Conclusions are given in Sec. 6.

2 Metasets

Metaset is a new approach to partial membership, similarly to fuzzy sets [17] and rough sets [9]. Metasets allow for representing imprecise notions. In this paper we focus on the vague idea of a *good car for a lady*. Members of this metaset are cars which match this idea to various degrees.

A metaset is a classical crisp set with a specific internal structure which encodes the membership degrees of its members. The membership degrees are expressed as nodes of the binary tree $\mathbb{T}$. All the possible membership values make up a Boolean algebra. They can be evaluated as real numbers. In applications we may use natural language terms for expressing the degrees.

2.1 Basic Definitions

A first-order metaset[4] is a relation between a set and a set of nodes of the binary tree $\mathbb{T}$.

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.

Thus, the structure we use to encode the degrees of membership is based on ordered pairs. The first element of each pair is the member and the second element is a node of the binary tree which contributes to the membership degree of the first element.

The binary tree $\mathbb{T}$ is the set of all finite binary sequences, 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 $\mathbb{1}$ being the empty sequence is the largest element of $\mathbb{T}$ in this ordering.

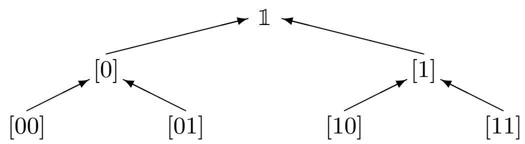

Fig. 1. The levels $\mathbb{T}_0$–$\mathbb{T}_2$ of the binary tree $\mathbb{T}$ and the ordering of nodes. Arrows point at the larger element.

We denote binary sequences which are elements of $\mathbb{T}$ using square brackets, for example: [00], [101]. If $p \in \mathbb{T}$, then we denote its children 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 $\mathbb{1}$ only. A *branch* in $\mathbb{T}$ is an infinite binary sequence. Abusing the notation we write $p \in \mathcal{C}$ to denote that the binary sequence $p \in \mathbb{T}$ is a prefix of the branch $\mathcal{C}$.

[4] For simplicity, in this paper we deal only with finite first-order metasets. See [11,12] for the introduction to metasets in general.

2.2 Interpretations

An interpretation of a first-order metaset is a crisp set. It is produced out of a given metaset using a branch of the binary tree. Different branches determine different interpretations of the metaset. All of them taken together make up a collection of sets with specific internal dependencies, which represents the source metaset by means of its crisp views. Properties of crisp sets which are interpretations of the given first-order metaset determine the properties of the metaset itself. In particular we use interpretations to define set-theoretic relations for metasets.

Definition 2. *Let τ be a first-order metaset and let $\mathcal{C}$ be a branch. The set*

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

is called the interpretation of the first-order metaset τ given by the branch $\mathcal{C}$.

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

The process of producing an interpretation of a first-order metaset consists in two stages. In the first stage we remove all the ordered pairs whose second elements are nodes which do not belong to the branch $\mathcal{C}$. The second stage replaces the remaining pairs – whose second elements lie on the branch $\mathcal{C}$ – with their first elements. As the result we obtain a crisp set contained in the domain of the metaset.

As we see, a first-order metaset may have multiple different interpretations – each branch in the tree determines one. Usually, most of them are pairwise equal, so the number of different interpretations is much less than the number of branches. Finite first-order metasets always have a finite number of different interpretations.

2.3 Partial Membership

We use interpretations for transferring set-theoretic relations from crisp sets onto metasets.[5] In this paper we discuss only the partial membership.

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

Formally, we define an infinite number of 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: $\sigma \epsilon_p \tau \wedge \sigma \epsilon_q \tau$. Membership under the root condition $\mathbb{1}$ resembles the full, unconditional membership of crisp sets, since it is independent of branches.

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

[5] For the detailed discussion of the relations or their evaluation the reader is referred to [12] or [14].

Example 1. Recall, that 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, $\sigma \epsilon_{\mathbb{1}} \tau$ holds.

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

2.4 Evaluating Membership

Membership degrees for metasets are expressed as nodes of $\mathbb{T}$. In fact, these nodes determine the basis of the Boolean Algebra of closed-open sets in the Cantor space 2^ω. Indeed, a $p \in \mathbb{T}$ is just a prefix for all infinite binary sequences which form a clopen subset of 2^ω. Thus, the membership relation for metasets is valued in the Boolean algebra. Nonetheless, for the sake of simplicity and in applications we usually refer to the binary tree when talking about membership.

In applications we frequently need a numerical evaluation of membership degrees. In order to define it, we first 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. Note, that by the definition 3, $\|\sigma \in \tau\| = \max\{p \in \mathbb{T} : \forall_{\mathcal{C} \ni p}\, \sigma_{\mathcal{C}} \in \tau_{\mathcal{C}}\}$. In other words, if $p \in \|\sigma \in \tau\|$, then for each branch $\mathcal{C}$ containing p holds $\sigma_{\mathcal{C}} \in \tau_{\mathcal{C}}$.

We define the numerical evaluation of membership taking the following assumptions. All 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 $\|\sigma \in \tau\| = \{\mathbb{1}\}$, i.e., $\sigma \epsilon_{\mathbb{1}}$ holds, then $|\sigma \in \tau| = 1$. And if $\|\sigma \in \tau\| = \emptyset$ ($\sigma \epsilon_p$ holds for no p), then $|\sigma \in \tau| = 0$.

For the sake of the main topic of the discussion it is worth stressing that in the above definition we treat all the nodes within the same level uniformly, without distinguishing one from another. This will not be the case for the problem of evaluation of client preferences, where we modify the above function to reflect interests in particular properties which compose an imprecise idea.

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 car for a lady.* Its members are particular cars 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. For instance, a good car must have good *looks* and be *comfortable.* But what does it mean to have good *looks*? For us, it means to have a nice *color* and *shape.* Similarly, we split the meaning of *comfortable* into sub-ideas. A *comfortable* car must have a friendly *user-interface* and must be fully *automated.* We might proceed splitting for arbitrary many steps. For the sake of simplicity we stop at the second step.

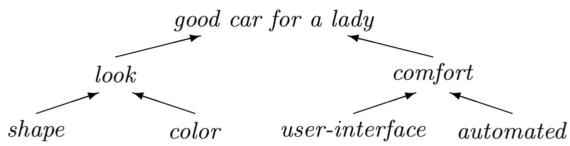

Fig. 2. The binary tree of the features describing a *good car for a lady*

The binary tree in the Fig. 2 is used throughout the paper to represent the discussed idea of a good car 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 car to satisfy the discussed idea.

3 Opinion Mining

Opinion Mining consists in identifying orientation or intensity of opinion in pieces of texts (blogs, forums, user comments, review websites, community websites, etc.). It enables determining whether a sentence or a document expresses positive, negative or neutral sentiment 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 $f_i \in F$ can be expressed with a finite set of words or phrases W_i, where W_i is a set of corresponding synonym sets $W_i = \{W_{i1}, W_{i2}, \ldots, W_{in}\}$ 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.

An opinion holder $j \in H$ makes comments concerning a subset of the features $S_j \subseteq F$ of an object O. For each feature $f_j \in Sj$ that the holder j comments on, the holder j chooses a word or phrase from W_k to describe the feature f_k, and expresses a positive, negative or neutral opinion on f_k.

In general, the first step of such a process is to retrieve the information from the Web [16] (tweets, blogs, forums, etc.) related to the object (O: *a good car for a lady* in our case, presented in Example 2), to extract the opinions about the selected features (F) and then to classify this information according to their emotional value.

Opinion Mining is a complex technique. Opinions can be expressed in a subtle manner which creates difficulty in the identification of their emotional values. Moreover, opinions are highly sensitive to the context and dependent of the field in which they are used: the same string might be positive in one context and negative in another. In addition, on the Internet, everyone uses his own style and vocabulary, that adds extra difficulty to the task. It is not yet possible to find out an ideal case to marking the opinion in a text written by different users, because the text does not follow the rules. Therefore, it is impossible to schedule every possible case. Moreover, very often the same phrase can be considered as positive for one person and negative for another one.

There are many methods used in Opinion Mining. We can divide the existing approaches in two categories: supervised and unsupervised methods. The most applied supervised learning techniques are Support Vector Machines and Naïve Bayes. These techniques give better results but at the same time they are very sensible to over-training and dependent on the quality, size and domain of the training data. The unsupervised approaches are based on external resources (dictionaries such as WordNet Affect or SentiWordNet, General Inquirer). The most painful disadvantages of these approaches are sensibility to the domain and dependence of the dictionary construction.

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

$$\text{SO-A}(word) = \Sigma_{p \in P} \, \mathrm{A}(word, p) - \Sigma_{n \in N} \, \mathrm{A}(word, n) \tag{1}$$

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.

To classify intensity of opinions concerning cars' buyers, we use the engine of our system [2,3,4].

4 Users' Profile Detection

Customer feedback is now targeted not only at companies directly, but also broadcast on the Net via weblogs, Twitter, Facebook, and comments at retailers' websites. This feedback can be very rich. It may consist of the evaluations of specific aspects of the product, information about the author/reviewer, and feedback from readers about the review, etc.

The objective of this section is to describe the framework to analyze the typical profile of a car buyer. The steps we take to achieve the goal are: selection of the features for the car, evaluation of their importance, retrieval and evaluation of the opinions for individual users, and finally – construction of particular user's profile. To perform these tasks we use Opinion Mining (OM) techniques.

In order to demonstrate our methodology we use the hierarchy of conditions depicted in the Fig. 2, which comprise the notion of a *good car for a lady.* The purpose is to find the opinions related to this topic. Particularly, for each holder $j \in H$ we want to find his/her opinions about the selected features F of cars and the intensities of these features, particularly, when the opinion holder buys the car. We consider that the opinions of both negative and positive polarity in the same way declare that the feature is important for the user. Therefore, the polarity of opinion (negative, positive) is not relevant, only the intensity of the opinion's value matters and it is considered to be the significant contribution to values of parameters. To calculate the intensity (SI – Semantic Intensity from Association) we use modified SO-A method:

$$\text{SI-A}(word) \quad = \quad \Sigma_{p \in P}\text{A}(word, p) \ + \ \Sigma_{n \in N}\text{A}(word, n)\ . \tag{2}$$

To find the intensity of each feature, we sort these opinions from highly-rated extremal opinion (positive, negative) to the neutral one.

By Def. 6, the basic components of an opinion are: object O (on which an opinion is expressed, in our case it is a car), opinion's holder j (a person that holds an specific opinion on a particular object), and sentiment/opinion (a view, attitude, or appraisal on an object from an opinion holder). According to the idea presented by the tree in the Fig. 2, the set F is composed of 6 elements $\{$*look, comfort, shape, color, user-interface, automated*$\}$.

For each element f_i we select manually the corresponding set w_i. For example,

$$w_{look} \quad = \quad \{\ \textit{appearance, outlook, aspect, air, outside,} \ldots \} \tag{3}$$

$$w_{comfort} \quad = \quad \{\ \textit{accommodation, commodious, convenience,} \ldots \} \tag{4}$$

We formalize the preferences of a sample customer in the following example.

Example 2. For demonstration purposes we consider here opinions of a sample user, named Ann, extracted from the Internet. Her posts included, among other pieces texts, also opinions of this sort: *"I love my new car, it's great, I can drive and call at the same time, my smartphone is connected"*, *"... when driving, I can listen to the music from my mp3"*, *"it had manual transmission, now I have automated one, but it didn't change a lot ..."*, *"it is red"*, *"its modern silhouette is very nice"*, *"the seats are really comfortable"*, and so on. Based on these we have estimated the following ratios for her preferences.

Ann prefers to have a comfortable car than a nice one. We found that her attitude is expressed by the ratio 3/2 in favor of *comfort* over *look*. She likes driving a lot, changing the transmission gears is not a problem for her. On the other hand, she likes to use her connected mobile devices a lot when she is driving, so the *user-interface* is very important feature for her. She professed a ratio of 3/1 in favor of *user-interface* over *automated*. According to her opinion some aspects of *shape* are critical for her, and therefore the discovered ratio is 4/1 over *color*.

5 Modeling with Metasets

We use the metaset approach to model the vague idea of a *good car for a lady*. Throughout the paper the metaset Δ represents the "fuzzy" collection of *good cars for a lady*. Potential members of Δ are the cars which match this imprecise notion more or less. Their membership degrees in Δ correspond to the levels of satisfaction of this property. When evaluating membership we assume that the capabilities of cars, represented by nodes within a level in $\mathbb{T}$, are equally important. By modifying the evaluation function so that some capabilities become more important than others, we may reflect particular clients' interests towards specific capabilities of cars which are Δ members. The real values obtained this way seem to reflect human reasoning. For instance, evaluating preference for a person interested in fast cars will result in higher value if a car has properties of a fast car indeed (acceleration, power), than when it has not.

5.1 Evaluating Client Preferences

Definition 5 (certainty value of membership) assumes uniform distribution of values throughout the nodes in $\mathbb{T}$: each $p \in \|\delta \in \Delta\|$ contributes the value of $\frac{1}{2^{|p|}}$ to $|\delta \in \Delta|$. In the context discussed in the paper this might be interpreted as a client's indifference as to what to choose: all possible choices represented as nodes within the same level are equally weighted. Particularly, for a $p \in \mathbb{T}$ both its children $p \cdot 0$ and $p \cdot 1$ contribute equally to the membership evaluation. In real life, however, clients have some preferences concerning choices they make. For instance, the *look* of the car might be more important than *comfort* for some clients. Such preferences, may be taken into account while evaluating client's attitude towards a particular instance of a car. We express these preferences numerically with the following function.

Definition 7. *We define client preference to be a function* $\mathfrak{p}\colon \mathbb{T} \mapsto [0,1]$ *such that*

$$\forall_{q\in\mathbb{T}} \;\; \mathfrak{p}(q\cdot 0) + \mathfrak{p}(q\cdot 1) = 1 \; . \tag{5}$$

and we take $\mathfrak{p}(\mathbb{1}) = 1$ *for the root.*

Given the preference function $\mathfrak{p}$ we evaluate the quality of a car δ taking preferences $\mathfrak{p}$ into account to obtain the subjective value of client's attitude towards δ. For this purpose we generalize the Def. 5 slightly to obtain an evaluation function which increases the impact of some nodes and decreases that of others. In applications we may (and we do in this paper) build this function based on the Opinion Mining techniques, as described in Sec. 4.

Definition 8. *Let* δ *be a car and let* Δ *be a metaset of good cars for a lady. The* $\mathfrak{p}$*-quality of the car* δ *is the following value:*

$$|\delta \in \Delta|_{\mathfrak{p}} = \sum_{q\in\|\delta\in\Delta\|} \; \prod_{0\leq i\leq |q|} \mathfrak{p}(q{\restriction_i}) \; .$$

The symbol $q{\restriction_i}$, where $0 \leq i \leq |q|$, denotes the prefix of the length i of the binary sequence q. For $i = 0$, it is the empty sequence $\mathbb{1}$, and for $i = |q|$, it is the q itself.

The $\mathfrak{p}$-quality of a car reflects client's preferences. For different clients with different $\mathfrak{p}$ preference functions it may result in different ratings for the given car.

Example 3. Based on the ratios discovered in the Ex. 2 we build the preference function $\mathfrak{p}$ for Ann (see Fig. 3). Since Ann prefers *comfort* to *look* with the ratio of 3/2, then we set $\mathfrak{p}(\textit{look}) = 0.4$ and $\mathfrak{p}(\textit{comfort}) = 0.6$. The ratio of 3/1 in favor of *user-interface* over *automated* results in setting $\mathfrak{p}(\textit{user-interface}) = 0.75$ and $\mathfrak{p}(\textit{automated}) = 0.25$. And since the ratio of *shape* over *color* is 4/1, then we have $\mathfrak{p}(\textit{shape}) = 0.8$ and $\mathfrak{p}(\textit{color}) = 0.2$.

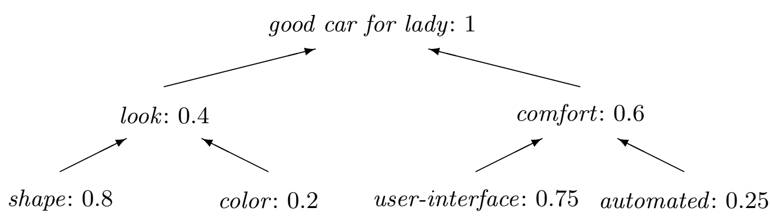

Fig. 3. The $\mathfrak{p}$ function for Ann.

5.2 Solution to the Problem

We use Ann's preferences described in the Examples 2 and 3 in order to demonstrate the mechanism of evaluating client's preferences and to show that indeed, it results in values reflecting human reasoning.

Let Δ be the metaset representing a *good car for lady*. Let α and β denote cars with the following capabilities. The α has perfect *shape* and it has a good *user-interface*, whereas β is fully *automated* and it has a nice *color*. Thus, we may write

$$\alpha \,\epsilon_{shape}\, \Delta \quad \wedge \quad \alpha \,\epsilon_{user\text{-}interface}\, \Delta \;, \tag{6}$$

$$\beta \,\epsilon_{automated}\, \Delta \quad \wedge \quad \beta \,\epsilon_{color}\, \Delta \;. \tag{7}$$

Therefore,

$$\|\alpha \in \Delta\| \quad = \quad \{\, shape, user\text{-}interface \,\} \quad = \quad \{\, [00], [10] \,\} \;, \tag{8}$$

$$\|\beta \in \Delta\| \quad = \quad \{\, color, automated \,\} \quad = \quad \{\, [01], [11] \,\} \;. \tag{9}$$

and

$$|\alpha \in \Delta| \quad = \quad \frac{1}{2^{|[00]|}} + \frac{1}{2^{|[10]|}} \quad = \quad \frac{1}{2^2} + \frac{1}{2^2} \quad = \quad 0.5 \;, \tag{10}$$

$$|\beta \in \Delta| \quad = \quad \frac{1}{2^{|[01]|}} + \frac{1}{2^{|[11]|}} \quad = \quad \frac{1}{2^2} + \frac{1}{2^2} \quad = \quad 0.5 \;. \tag{11}$$

We see, that both cars satisfy the requirements for the *good car for a lady* to the same degree of 0.5.

However, if we take into account the client's preferences expressed as the preference function $\mathfrak{p}$, then these cars turn out to be quite different.

$$|\alpha \in \Delta|_{\mathfrak{p}} \quad = \quad 0.4 \cdot 0.8 + 0.6 \cdot 0.75 = 0.77 \;, \tag{12}$$

$$|\beta \in \Delta|_{\mathfrak{p}} \quad = \quad 0.4 \cdot 0.2 + 0.6 \cdot 0.25 = 0.23 \;. \tag{13}$$

We conclude, that Ann's interest in the car α is much greater than in car β. The values and the relation $|\alpha \in \Delta|_{\mathfrak{p}} > |\beta \in \Delta|_{\mathfrak{p}}$ confirm her opinions shared on the Internet, as described in the Ex. 2.

6 Conclusions

The aim of our research is to design a decision support system which can find its application in traditional marketing and e-marketing. In this paper we show how methods and techniques of Opinion Mining can be used for selecting from users' declarations the most important features of products as well as for building users' profiles. This data is represented as a metaset and then analyzed. In this way, the theory of metasets is applied for evaluating clients' preferences. Our approach is used in designing a software tool which supports making decisions. It can help marketers evaluate the success of an ad campaign or a new product and identify the product features which the users like or dislike.

Acknowledgments Kacprzak thanks the Polish Ministry of Science and Higher Education at the Bialystok University of Technology under 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)
2. Dziczkowski, G., Wegrzyn-Wolska, K.: Rcss - rating critics support system purpose built for movies recommendation. In: Advances in Intelligent Web Mastering. Springer (2007)
3. 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. Esuli, A., Sebastiani, F.: SentiWN: A Publicly Available Lexical Resource for Opinion Mining. Proc. of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), 417-422 (2006)
6. Jech, T.: Set Theory: The Third Millennium Edition, Revised and Expanded. Springer-Verlag, Berlin Heidelberg New York (2006)
7. Kacprzak, M., Starosta B.: An Approach to Making Decisions with Metasets. In Trends in Contemporary Computer Science, Podlasie 2014, by A. Gomolinska, A. Grabowski, M. Hryniewicka, M. Kacprzak, E. Schmeidel (Eds.), 159–172 (2014)
8. Kunen, K.: Set Theory, An Introduction to Independence Proofs. Studies in Logic and Foundations of Mathematics, North-Holland Publishing Company (1980)
9. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences 11, 341–356 (1982).
10. Starosta, B.: Application of Metasets to Character Recognition. In: Rauch, J. et al. (eds.), ISMIS 2009. LNCS vol. 5722, pp. 602–611. Springer-Verlag (2009)
11. Starosta, B.: Metasets: A New Approach to Partial Membership, In: Rutkowski, L. et al. (eds.), Artificial Intelligence and Soft Computing, LNAI 7267 pp. 325–333. Springer-Verlag (2012)
12. 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. Springer-Verlag Berlin Heidelberg (2009)
13. Starosta, B., Kosiński, W.: Metasets, Intuitionistic Fuzzy Sets and Uncertainty. In: Rutkowski, L. et al. (eds.), Artificial Intelligence and Soft Computing, LNAI 7894 pp. 388–399. Springer-Verlag (2013)
14. 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. Systems Research Institute, Polish Academy of Sciences, Warsaw (2013).
15. Souffriau, W., Vansteenwegen P., Vertommen J., Vanden Berghe G., Van Oudheusden D.: A personalizes tourist trip design algorithm for mobile tourist guides, Applied Artificial Intelligence, vol. 22(10) pp. 964-985 (2008)
16. Wegrzyn-Wolska, K., Bougueroua, L.: Tweets mining for French Presidential Election, In Proc. of the 4th IEEE/WIC International Conference on Computation Aspects of Social Networks - CASoN 2012, SaO Carlos, Brazil, November (2012)
17. Zadeh, L. A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)