A highly automated recommender system based on a possibilistic interpretation of a sentiment analysis

Abdelhak Imoussaten, Benjamin Duthil, François Trousset, Jacky Montmain

A Highly Automated Recommender System Based on a Possibilistic Interpretation of a Sentiment Analysis

Abdelhak Imoussaten[1,⋆], Benjamin Duthil[2], François Trousset[1], and Jacky Montmain[1]

[1] LGI2P, Ecole des Mines d'Alès,
Site EERIE Parc Scientifique G. Besse, 30035 Nîmes Cedex, France
[2] L3i, Université de La Rochelle,
Avenue M. Cré peau, 17042 La Rochelle cedex 01, France
{abdelhak.imoussaten,francois.trousset,jacky.montmain}@mines-ales.f,
benjamin.duthil@univ-lr.fr

Abstract. This paper proposes an original recommender system (RS)[1] based upon an automatic extraction of trends from opinions and a multicriteria multi actors assessment model. Our RS tries to optimize the use of the available information on the web to reduce as much as possible the complex and tedious steps for multicriteria assessing and for identifying users' preference models. It may be applied as soon as i) overall assessments of competing entities are provided by trade magazines and ii) web users' critics in natural languages and related to some characteristics of the assessed entities are available. Recommendation is then based on the capacity of the RS to associate a web user with a trade magazine that conveys the same values as the user and thus represents a reliable personalized source of information. Possibility theory is used to take account subjectivity of critics. Finally a case study concerning movie recommendations is presented.

Keywords: Possibility theory, Intervals merging, Multicriteria aggregation, Recommender system, Opinion-mining.

1 Introduction

In recent years, many companies and web sites have set up systems to analyze the preferences of their users in order to better meet their expectations. To date, recommendation systems are present in many areas such as tourism /leisure, advertising, e-commerce, movies, etc. Due to the exponential growth of the quantity of data available on the Internet in recent years, searching and finding products, services and relevant contents become a difficult task for the user often drowned out by the mass of information. This explains the growing interest in recommendation systems (RS) both by users as by commercial sites.

⋆ Corresponding author.

The task of recommendation has been identified as a way to help users to find information, or elements that are likely of interest. Roughly speaking, we consider a set of users and a set of items (products or services) that may be recommended to each user. In addition, a multicriteria recommendation improves the quality of a RS because it makes explicit the characteristics for which an item was proposed to the user [2] [3]. A RS that takes advantage of evaluation related to multicriteria preference elicitation provides users with more relevant detailed recommendations [2]. However, the implementation of such a model requires a knowledge base where the items are evaluated *w.r.t* a set of criteria. This constraint imposed by the model is very heavy for the user.

In [4], an unsupervised multicriteria opinion mining method is proposed. It allows users to free themselves from constraining partial evaluations *w.r.t* each criterion: users simply submit their critics in natural language to express their opinions, and system analyzes them automatically. It first dissects the critics according to the evaluation criteria (thematic segmentation) before calculating the polarity or opinions of each of the extracts resulting from the segmentation step (opinion-mining/sentiment analysis).

Combining this method with an interactive multicriteria decision support system makes possible to have a highly automated system for recommendation purposes. However, the automated assignment only provides imprecise scores related to items that are modeled by intervals in an adequate multicriteria analysis process. Our possibility theory based approach then manages multiple imprecise assessments derived from sentiment analysis on each evaluation criterion (intervals fusion) and then aggregate them on all criteria. Finally, we try to match a user and an adequate specialized magazine in the domain of concern (movies in our application) that will provide the most suitable personalized recommendation to the user.

Section 2 summarizes opinion mining approach to extract Internet user's critics and compute opinion scores on a set of criteria. Section 3 explains how to merge these imprecise opinion scores for each criterion and introduces the notion of matching of a distribution with data available on a criterion. Section 4 describes how to deduce the multicriteria model used by a specialized magazine to assess items. Section 5 shows how these approaches can be combined to address the multicriteria recommendation problem in the case study of movie recommendations.

2 Opinion-Mining Process

On the basis of statistical methods, the Opinion-mining approach of [4] allows us to build a lexicon of opinion descriptors for a given thematic. This lexicon is used to automatically extract the polarity of text segments that are related to the criterion. Two stages are distinguished in this multicriteria evaluation: - firstly, the segments of text related to each of the evaluation criteria are extracted with the *Synopsis* approach described in [5]. The text is first segmented into criteria. Then, for each criterion, the polarities of the segments that have been identified

by the *Synopsis* approach are computed. This is opinion-mining or sentiment analysis step.

2.1 Text Segmentation by Criterion

The Synopsis approach is used to identify the text extracts that refer to the adopted criteria and consists of 3 steps:

1. automatic construction of a training corpus used to learn characteristic words (or groups of words), called *descriptors*, for a criterion of interest
2. Automatic learning of these *descriptors* and construction of a lexicon associated with the criterion
3. Text segmentation using the lexicon associated with the criterion

The approach uses a set of *seed words*. On the one hand, they serve to semantically characterize the criterion of concern, and on the other hand to initiate the learning of *descriptors* of the criterion [5]. The training corpus is built automatically. The hypothesis of the learning is based on the fact that the more frequently a descriptor is found in the neighborhood of a *seed word* of the criterion (counting on sliding window), the greater the membership function of this descriptor to the lexical scope of the criterion.

Synopsis builds automatically its training corpus for each criterion by using web documents. This corpus is used to automatically build a lexicon of descriptors for each of these criteria. Each lexicon is then used by the segmentation process to automatically extracts parts of text dealing with this criterion. Synopsis will identify several segmentation for each criterion depending of granularity levels. Those granularities can be associated with user's level of expertise. The pertinent levels are automatically identified by Synopsis (see [4] and [5] for more details).

2.2 Opinion Analysis

The opinions extraction approach is an adaptation of the *Synopsis* approach to the extraction of opinions both for the automatic building of the training corpus and for the descriptors learning phases. Seed words become opinion seed words [4]. Two sets of words of opinion are initially distinguished the positive ones: $P = \{good, nice, excellent, positive, fortunate, correct, superior\}$; and the negative ones: $N = \{bad, nasty, poor, negative, unfortunate, wrong, inferior\}$.

Assuming a document that contains at least one word of P (resp. N) and none of N (resp. P) conveys a positive opinion (resp. negative), a set of documents associated with a seed word is built with a search engine on the web as in *Synopsis* (request for seed word "good" in the movie domains: movie + good - bad - nasty - negative - poor - unfortunate - wrong - inferior). Opinion descriptors are limited to adjectives and adjectival groups [4]. Statistical techniques of filtering with sliding windows of *Synopsis* are adapted to count opinion descriptors occurences.

Finally, an opinion score for a criterion is provided as a weighted sum of the text segments' membership related to the criteria and the positive/negative

descriptors. Thus, since the extraction of criteria segments depends on the level of expected precision (*i.e.* level of expertise), the score of the opinion text related to the criterion is also affected by the discrimination threshold. For each threshold the segmentation algorithm of [4] generates the corresponding text segmentation. Then the opinion scores of the text can be computed for any user's expertise level. There exists a lower and upper bound for this score. Accordingly, the opinion score is an imprecise entity which is then represented as an interval whose extremities are these lower and upper bounds.

3 Intervals Merging

The imprecision involved here concerns the subjectivity of a critic in the evaluation process of a text. This "subjectivity" is technically related to the imprecision of extraction. It is however "homogenized" by the automatic processing of segmentation. Then, evaluation is also uncertain because of the multiplicity of automatically collected opinions. Belief theory [6], [7] provides an appropriate framework to summarize these opinions and ease their representation and manipulation in the recommendation process while respecting their imprecision and uncertainty. Possibility distributions [8] are good approximations of belief functions. Thus they will be used to represent assessments. Also, possibility functions are appealing from an interpretation point of view in collecting confidence intervals as well as a computational point of view.

3.1 Possibility Theory

Let Ω represent a universal set of elements ω under consideration that is assumed to be finite and let 2^Ω represent the power set of Ω. A possibility distribution π is a normalized function $\pi : \Omega \longrightarrow [0,1]$ (i.e. $\exists \omega \in \Omega$, such that $\pi(\omega) = 1$). From π, possibility and necessity measures are respectively defined for all subsets $A \in 2^\Omega$: $\Pi(A) = \sup_{\omega \in A} \pi(\omega)$ and $N(A) = 1 - \Pi(A^c)$. $\Pi(A)$ quantifies to what extent the event A is plausible while $N(A)$ quantifies the certainty of A. An α-cut of possibility distribution π is the classical subset: $E_\alpha = \{\omega \in \Omega : \pi(\omega) \geq \alpha\}$, $\alpha \in]0,1]$. When a distribution has a trapezoidal form, it is classically represented by its vertices abcd.

3.2 Evidence Theory

The evidence theory shall now be formulated by the basic belief assignment (*bba*) m defined from 2^Ω to $[0,1]$, such that: $\sum_{\{A \subseteq \Omega\}} m(A) = 1$ and $m(\emptyset) = 0$. Elements E of 2^Ω such that $m(E) > 0$ are called focal elements and their set is denoted $\mathtt{F}$. The *bba* m can be represented by two measures: the belief function $Bel(A) = \sum_{\{E \in \mathtt{F}/A \supseteq E\}} m(A)$, $A \subseteq \Omega$ and the plausibility function $Pl(A) = \sum_{\{E \in \mathtt{F}/A \cap E \neq \emptyset\}} m(A)$, $A \subseteq \Omega$. When focal elements are imprecise, the probability of any event $A \subseteq \Omega$, denoted $Pr(A)$, is imprecise and $Bel(A)$ and

$Pl(A)$ represent respectively the lower and upper probabilities of event A, *i.e.* $Pr(A) \in [Bel(A), Pl(A)]$. Two well-known extreme cases of belief and plausibility measures are probability measures and possibility measures [8].

3.3 Building Possibility Distributions From a Set of Intervals

Let consider a set of distinct intervals $\{I_j, j = 1, nbi\}$ as the focal elements and the probability of occurrence of interval I_j as the *bba* $m(I_j)$ assigned to this interval. When intervals are nested, *i.e.* $I_1 \subset I_2 \subset ... \subset I_{nbi}$, a possibility distribution π may be built from plausibility measure, as proposed in [9]: $\forall \omega \in \Omega$, $\pi(\omega) = Pl(\{\omega\}) = \sum_{j=1,nbi} m(I_j).\mathbf{1}_{I_j}(\omega)$. When intervals are consistent, *i.e.* $\bigcap_{j=1,nbi} I_j = I \neq \emptyset$ (all experts share at least one value), but not nested, two possibility distributions π_1 and π_2 are built: First, we consider the *bba* $m_1(I_j)$ for focal elements $\{I_j, j = 1, nbi\}$. Thus $\forall \omega \in \Omega$, $\pi_1(\omega) = \sum_{j=1,nbi} m_1(I_j).\mathbf{1}_{I_j}(\omega)$. Second, r nested focal elements $\{E_s, s = 1, r\}$ are obtained from original data from the α-cuts of π_1: $E_1 = I$ and $E_s = E_{s-1} \cup E_{\alpha_s}(\pi_1)$ $(s = 2, r)$. The new bba m_2 assigned to intervals E_s are computed as proposed in [9]: $m_2(E_s) = \sum_{\{I_j \ related \ to \ E_s\}} m_1(I_j)$ (each assessments I_j being *related* in a unique way to the smallest E_s containing it). Then a possibility distribution π_2 can be defined as: $\forall \omega \in \Omega$, $\pi_2(\omega) = \sum_{s=1,r} m_2(E_s).\mathbf{1}_{E_s}(\omega)$. Membership functions π_1 and π_2 are mono modal possibility distributions since $\bigcap_{j=1,nbi} I_j = I \neq \emptyset$ holds. Furthermore, they are the best possibilistic lower and upper approximations (in the sense of inclusion) of assessment sets $\{I_j, j = 1, nbi\}$ [9]. It can be seen easily that $\pi_1 \subseteq \pi_2$ (inclusion of fuzzy subsets) as $\forall \alpha \in]0, 1]$, $E_{1,\alpha} \subseteq E_{2,\alpha}$.

In general however, experts' assessment might be neither precise nor consistent. The probability and possibility representations correspond respectively to extreme and ideal situations; unfortunately, critics may reveal contradictory assessments. This means that the consistency constraint may not be satisfied in practice, i.e. $\bigcap_{j=1,nbi} I_j = \emptyset$. To cope with this situation, groups of intervals, *maximal coherent subsets* (*MCS*), with a non-empty intersection are built from original intervals, which is equivalent to find subsets $K_\beta \subset \{1, ..., nbi\}$ with $\beta \in \{1, ..., g\}$ such that: $\bigcap_{j \in K_\beta} I_j \neq \emptyset$, with g being the number of subsets K_β [10].

For each group K_β, lower and upper possibilistic distributions π_1^β and π_2^β are built (as in the previous case when elements are consistent). Let possibility distribution π_1 (resp. π_2) be the union (denoted $\tilde{\bigcup}$) of possibility distributions π_1^β (resp. π_2^β):

$$\pi_1 = \tilde{\bigcup_{\beta=1,g}} \pi_1^\beta \ (resp. \ \pi_2 = \tilde{\bigcup_{\beta=1,g}} \pi_2^\beta) \tag{1}$$

then π_1 and π_2 are the multi-modal (g modes) possibilistic lower and upper approximations of original intervals.

Reasoning with the lower distribution (resp. upper distribution) might correspond to a severe risk aversion position relative to the probability of information (resp. a flexible risk acceptance position). To maintain the richness of information provided by critics, the best way is to keep both distributions.

3.4 Matching between Distribution and a Set of Intervals

Let consider possibility distributions π_1 and π_2 the possibilistic approximations of intervals $\{I_j, j = 1, nbi\}$. Let denote for π_1 (resp. π_2) (N_1, Π_1) (resp. (N_2, Π_2)) the associated possibility and necessity measures. Then π_1 and π_2 are respectively the *greatest* and *smallest* fuzzy subsets such that [9]: $\forall A \subseteq \Omega, [N_1(A), \Pi_1(A)] \subseteq [Bel(A), Pl(A)] \subseteq [N_2(A), \Pi_2(A)]$. Let still consider two possibility distributions π and π^* defined on Ω. Two definitions are introduced:

Definition 1. *We define the degree of inclusion of π in π^* as:*

$$incl(\pi, \pi^*) = (\int_\Omega (\pi^* \wedge \pi))/\int_\Omega \pi \tag{2}$$

Definition 2. *We define the degree of matching of π to data $\{I_j, j = 1, nbi\}$ as:*

$$match(\pi, \{I_j\}) = [incl(\pi_1, \pi) + incl(\pi, \pi_2)]/2 \tag{3}$$

$\{I_j\}$ is said to match π better than π^* if: $match(\pi^*, \{I_j\}) < match(\pi, \{I_j\})$. We also use the notation $match(\pi, (\pi_1, \pi_2))$ instead of $match(\pi, \{I_j\})$ when possible.

Remark 1. Definition 2 leads, for particular cases of π, to:

- $match(\pi, \{I_j\}) \in [0, 1]$.
- If $\pi_1 \subseteq \pi \subseteq \pi_2$ (inclusion of fuzzy subsets), then $match(\pi, \{I_j\}) = 1$.
- If $\pi_2 \subset \pi$, then $match(\pi, \{I_j\}) = [1 + (\int_\Omega \pi_2 / \int_\Omega \pi)]/2$.
- If $\pi \subset \pi_1$, then $match(\pi, \{I_j\}) = [(\int_\Omega \pi / \int_\Omega \pi_1) + 1]/2$.
- If $\pi_2 \cap \pi = \emptyset$, then $match(\pi, \{I_j\}) = 0$.

The idea behind the *matching* is to consider that the distribution π (with N and Π its associated possibility and necessity measures) which guarantees $[N_1(A), \Pi_1(A)] \subseteq [N(A), \Pi(A)] \subseteq [N_2(A), \Pi_2(A)]$ for a large number of subset $A \subseteq \Omega$, *better* matches data $\{I_j, j = 1, nbi\}$. This approach is similar to the one in [11] which used in fuzzy pattern matching.

4 Identification of Preferences Model

Aggregation models make the capture of the notion of priorities in the decision-maker's strategy possible, and simplify the comparison of any two alternatives described through their elementary evaluation. The most commonly used operator to express decision maker preferences is the weighted average mean denoted

here WAM_ω. It allows giving non-symmetrical roles to criteria through a vector of relative weights $\omega = (\omega_1, ..., \omega_n) \in [0,1]^n$.

The specialized press generally provides a simple overall assessment of items using evaluation scales such as: number of stars, number of bars, etc.. This score is accompanied by a more or less detailed critic in natural language which is supposed to make explicit the assessment of the journal. However it is often difficult for users to understand the exact reasons that would justify the imprecise overall score the item received. We model these imprecise scores by possibility distributions $\widetilde{\pi}^k$ on $[0, 20]$ where k is an item index. To make clearer assessments allocated by a magazine to an item, we try to identify the strategy, the priorities that characterize this journal: what are the criteria that differentiate the values conveyed by this journal? As soon as a user knows which critics are of interest in the assessments of a specialized magazine, he can then choose the journal that matches the best his priorities and choose a film recommended by this magazine.

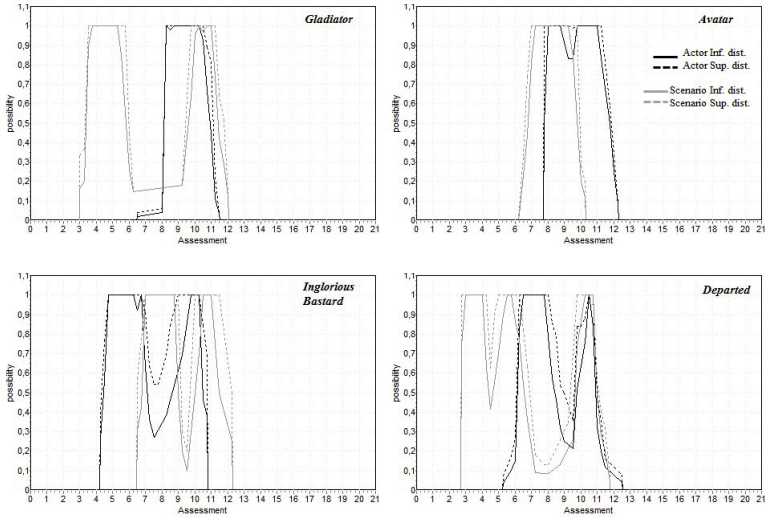
Many Internet users contribute to collaborative recommender systems but provide their opinion in natural language because they are not familiar with assessments and all the less with multi criteria assessments. Our segmentation and sentiment analysis system automatically collects all these criteria and, for each critic c that deals with an identified criterion i, it extracts an imprecise score (depending of the analysis granularity, *i.e.* the expected level of expertise) which is modeled as an interval as explained in section 2.2. According to the merging method presented in section 3.3 we compute two possibility distributions $\pi^k_{i,1}$ (inferior) and $\pi^k_{i,2}$ (superior) for the set of automatically collected imprecise scores for each item k and each criterion i. Let us note $\overline{\pi_{\alpha,\omega}}^k = WAM_\omega(\pi^k_{\alpha,1}, ..., \pi^k_{\alpha,n})$ with $\alpha \in \{1, 2\}$. The next step is to identify the weights of the distribution that best match the magazine's overall assessments. In other words, we search for weights ω such that $\widetilde{\pi}^k$ and $(\overline{\pi_{1,\omega}}^k, \overline{\pi_{2,\omega}}^k)$ match as well as possible for a learning set of items (items). Mathematically, a possible answer is based on our function *match* defined in section 3.4 such as:

$$\omega^* = Arg \max_{\omega \in [0,1]^n} \sum_{k \in items} match(\widetilde{\pi}^k, (\overline{\pi_{1,\omega}}^k, \overline{\pi_{2,\omega}}^k)) \quad (4)$$

5 Case Study

The software prototype that supports our recommendation system is based on a combination of an Internet user with a specialized journal that bears the same priorities or values as him. The case study presented in this section is based on movie recommendations. This prototype uses the multicriteria opinion extraction module in section 2, as well as, a base of movie critics written in natural language from the famous film critics site IMDB. Critics provided by IMDB (about 3000 critics per movie) provide enough information to get a representative picture of the diversity of opinions about this movie. Each film critic has been evaluated by our multicriteria opinions extraction system. To illustrate the method and simplify the presentation only two criteria are considered in this toy case study:

Table 1. Inferior and superior possibility distributions



actor and *scenario*. With the merger process of Section 3.3, we obtain inferior and superior distributions for 20 films in this case study. Table 1 shows the results respectively for movies: Gladiator, Avatar, Inglorious Bastard and Departed.

As we can see in table 1 multi modality is present in all distributions. It is due to the divergence found in critics' opinions: users are rarely unanimous on a movie review.

We selected 39 journals in this application. For each movies of the base, some journals provide overall assessments in the form of a number of stars (transformed into trapezoidal distributions $\widetilde{\pi}^k$) (see table 5). The weight distribution that characterize the best the journal evaluation strategy is calculated through equation 4 for each journal: *e.g.* the weight distribution that explains at best the scores $\widetilde{\pi}^k$ assigned to the 39 movies by the journal in the learning database. Fig. 1 shows that there are large differences between assessment strategy of journals (*e.g.* for Cahiers du Cinema, the weights are 0.6 for *actor* and 0.4 for *scenario* while conversely they are 0.22 and 0.78 for Charlie Hebdo). Some journals attach no importance to actors and some others to scenario.

Note that for a reliable and relevant recommendation, our model should integrate more criteria in the assessment. Finally, Fig. 1 provides to the Internet user how important criteria are considered in the 39 journals assessment strategy. He

Table 2. Transformation of stars to trapezoidal distributions

numbers of stars	*/2	*	*(*/2)	...	****
abcd trapezium	{0,2.5,5,7.5}	{2.5,5,7.5,10}	{5,7.5,10,12.5}	...	{12.5,15,17.5,20}

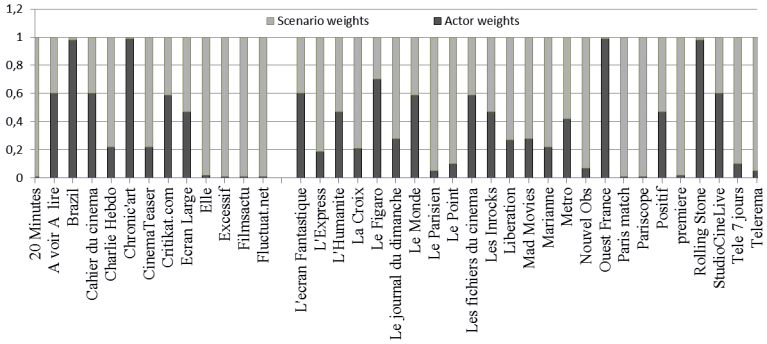


Fig. 1. Assessment strategy of movies specialized journals

may then simply choose the journal that conveys the values that are closest to his mood of the moment and consult the hit list of this journal for a personalized recommendation. No model of user preference is need to be identified which is generally a thorny and forbidding task in collaborative RS systems. The user identifies himself the journal that suits him as *best*. Note that this simple principle allows the user to change his "preferences system" depending on his mood every time he goes to the cinema!

6 Conclusion

The automated extraction of critics related to a set of criteria, the imprecise assessment process based on our sentiment analysis and our fuzzy multicriteria analysis allows the development of highly automated recommender systems of type "multicriteria preference elicitation from evaluations"[2] free of the most constraining tasks of this type of collaborative systems. This is an important step because until now the need to manually assess a large number of documents according to several criteria represented a major obstacle to the implementation of such systems. The process we propose establishes a cognitive automation that can be easily deployed into Web applications.

References

1. Imoussaten, A., Duthil, B., Trousset, F., Montmain, J.: Possibilistic interpretation of a sentiment analysis for a web recommender system. In: LFA 2013, Reims (2013)
2. Adomavicius, G., Manouselis, N., Kwon, Y.: Multi-criteria recommender systems. In: Recommender Systems Handbook, pp. 769–803. Springer (2011)
3. Manouselis, N., Costopoulou, C.: Analysis and classification of multi-criteria recommender systems. World Wide Web 10(4), 415–441 (2007)
4. Duthil, B., Trousset, F., Dray, G., Montmain, J., Poncelet, P.: Opinion extraction applied to criteria. In: Liddle, S.W., Schewe, K.-D., Tjoa, A.M., Zhou, X. (eds.) DEXA 2012, Part II. LNCS, vol. 7447, pp. 489–496. Springer, Heidelberg (2012)

5. Duthil, B., Trousset, F., Roche, M., Dray, G., Plantié, M., Montmain, J., Poncelet, P.: Towards an automatic characterization of criteria. In: Hameurlain, A., Liddle, S.W., Schewe, K.-D., Zhou, X. (eds.) DEXA 2011, Part I. LNCS, vol. 6860, pp. 457–465. Springer, Heidelberg (2011)
6. Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Annals of Mathematical Statistics 38 (1967)
7. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
8. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
9. Dubois, D., Prade, H.: Fuzzy sets and statistical data. European Journal of Operational Research 25 (1986)
10. Imoussaten, A., Mauris, G., Montmain, J.: A multicriteria decision support system using a possibility representation for managing inconsistent assessments of experts involved in emergency situations. IJIS 29(1), 50–83 (2014)
11. Dubois, D., Prade, H., Testemale, C.: Weighted fuzzy pattern matching. Fuzzy Sets and Systems 28(3), 313–331 (1988)