

Inductive Classification Through Evidence-Based Models and Their Ensembles

Giuseppe Rizzo^(✉), Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito

LACAM – Dipartimento di Informatica, Università degli Studi di Bari
“Aldo Moro”, Bari, Italy

{giuseppe.rizzo,claudia.damato,nicola.fanizzi,floriana.esposito}@uniba.it

Abstract. In the context of Semantic Web, one of the most important issues related to the class-membership prediction task (through inductive models) on ontological knowledge bases concerns the imbalance of the training examples distribution, mostly due to the heterogeneous nature and the incompleteness of the knowledge bases. An ensemble learning approach has been proposed to cope with this problem. However, the majority voting procedure, exploited for deciding the membership, does not consider explicitly the uncertainty and the conflict among the classifiers of an ensemble model. Moving from this observation, we propose to integrate the Dempster-Shafer (DS) theory with ensemble learning. Specifically, we propose an algorithm for learning *Evidential Terminological Random Forest* models, an extension of Terminological Random Forests along with the DS theory. An empirical evaluation showed that: (i) the resulting models performs better for datasets with a lot of positive and negative examples and have a less conservative behavior than the voting-based forests; (ii) the new extension decreases the variance of the results.

1 Introduction

In the context of Semantic Web (SW), ontologies and the ability to perform reasoning on them, via deductive methods, play a key role. However, standards inference mechanisms have also shown their limitations due to the incompleteness of ontological knowledge bases deriving from the Open World Assumption (OWA). In order to overcome this problem, alternative forms of reasoning, such as *inductive reasoning*, have been adopted to perform various tasks such as concept retrieval and query answering [1, 2]. These tasks have been cast as a classification problem, consisting in deciding the class-membership of an individual with respect to a query concept, to be solved through inductive learning methods that exploit statistical regularities in a knowledge base. The resulting models can be directly applied to the knowledge base or mixed with deductive reasoning capabilities [3]. Although the application of these methods has shown interesting results and the ability to induce assertional knowledge that is not logically derivable, these methods have also revealed some problems due to the aforementioned incompleteness. In general, the individuals that are positive and negative instances for a given concept may not be equally distributed. This skewness may be stronger when considering individuals whose membership cannot

be assessed because of the OWA. This *class-imbalance* setting may affect the model, resulting with poor performances.

Various methods have been devised for tackling the problem, spanning from sampling methods to ensemble learning approaches [4]. Concerning the specific task of instance classification for inductive query answering on SW knowledge bases, we investigated on the usage of ensemble methods [5], where the resulting model is built by training a certain number of classifiers, called *weak learners*, and the predictions returned by each weak learner are combined by a rule standing for the *meta-learner*. Specifically, we proposed an algorithm for inducing *Terminological Random Forests* (TRFs) [5], an ensemble of *Terminological Decision Trees* (TDTs) [6]. The method extends *Random Forests* and *First Order Random Forests* [7, 8] to the case of DL representation languages. When these models are employed, the membership for a test individual is decided according to a majority vote rule (although various strategies for combining predictions have been proposed [9–11]): each classifier returning a vote in favor of a class equally contributes to the final decision. In this way, some aspects are not considered explicitly, such as the uncertainty about the class label assignment and the disagreement that may exist among weak learners. The latter plays a crucial role for the performance of ensemble models [12]. In the specific case of TRFs, we noted that most misclassifications were related to those situations in which votes are distributed evenly with respect to the admissible labels.

A weighted voting procedure may be an alternative strategy to mitigate the problem, but it requires a criterion for setting the weights. In this sense, introducing a meta-learner which manipulates *soft predictions* of each classifier (i.e. a prediction with a confidence measure for each class value) rather than *hard predictions* (where a class value is returned) may be a solution. For TRFs, this can be done by considering the extension of TDT models based on the *Dempster-Shafer Theory* (DS) [13], which provides an explicit representation of ignorance and uncertainty (differently from the original version proposed in [6]). In machine learning, resorting to the DS operators is a well-known solution [14]. Most of the existing ensemble combination methods resort to a solution based on *decision templates*, which are obtained by organizing, for each classifier against each class, a mean vector (called *reference vector*). When these methods are employed, predictions are typically made by computing the similarity value between a decision profile of an unknown instance with the decision templates. Other approaches that does not require the computation of these matrices have been proposed [14]. However, all the methods consider a propositional representation. Additionally, none of them has been employed for predicting assertions on ontological knowledge bases.

The main contribution of the paper concerns the definition of a framework for the induction of *Evidential Terminological Random Forests* for ontological knowledge bases. This is an ensemble learning approach that employs Evidential TDTs (ETDTs) [13] and does not require the computation of decision templates, similarly to [14]. After the induction of the forest, a new individual is classified by combining, by means of the Dempster's rule [15], the available evidence on the membership coming from each tree.

The remainder of the paper is organized as follows: the next section recalls the basics of the Dempster-Shafer Theory; Sect. 3 presents the novel framework for evidential terminological random forests, while in Sect. 4, a preliminary empirical evaluation is described. Sect. 5 draws conclusions and illustrate perspectives for further developments.

2 Basics on the Dempster-Shafer Theory

The Dempster-Shafer Theory (DS) is basically an extension of the Bayesian subjective probability. In the DS, the *frame of discernment* is a set of exhaustive and mutually exclusive hypotheses $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ about a domain. For instance, the frame of discernment for a classification problem could be the set of all admissible class values. Moving from this set, it is possible to define a Basic Belief Assignment (BBA) as follows:

Definition 1 (Basic Belief Assignment). *Given a frame of discernment $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$. A Basic Belief Assignment (BBA) is a function that defines a mapping $m : 2^\Omega \rightarrow [0, 1]$ such that:*

$$\sum_{A \in 2^\Omega} m(A) = 1 \tag{1}$$

Given a piece of evidence, the value of a BBA m for a set A expresses a measure of belief exactly committed to A . This means that the value $m(A)$ does imply no further claims about any of its subsets. This means that when $A = \Omega$, a case of *total ignorance* occurs. Each element $A \in 2^\Omega$ for which $m(A) > 0$ is said to be a *focal element* for m . The function m can be used to define other functions, such as the belief and the plausibility function.

Definition 2 (Belief Function and Plausibility Function). *For a set $A \subseteq \Omega$, the belief in A , denoted $Bel(A)$, represents a measure of the total belief committed to A given the available evidence.*

$$\forall A, B \in 2^\Omega \quad Bel(A) = \sum_{B \subseteq A} m(B) \tag{2}$$

The plausibility of A , denoted $Pl(A)$, represents the amount of belief that could be placed in A , if further information became available.

$$\forall A, B \in 2^\Omega \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{3}$$

It can be proved that, knowing just one among m , Bel and Pl allows to derive all the other functions [16].

In the DS, various measures for quantifying the amount of uncertainty have been proposed, e.g. the *non-specificity measure* [17]. The latter can be regarded

as a measure for representing the imprecision of a BBA function. This measure can be computed by the following equation:

$$Ns = \sum_{A \in 2^\Omega} m(A) \log(|A|) \tag{4}$$

It is easy to note that the non-specificity value is higher when the focal elements are larger subsets of Ω , for the elements of which no further claims can be made.

One of the most important aspects related to the DS is the availability of various operators for pooling evidence from different sources of information. One of them, called *Dempster’s rule*, aggregates independent evidences defined within the same frame of discernment. Let m_1 and m_2 be two BBAs. The new BBA obtained by combining m_1 and m_2 using the rule of combination, m_{12} , can be expressed by the orthogonal sum of m_1 and m_2 . Generally, the normalized version of the rule is used:

$$\forall A, B, C \subseteq \Omega \quad m_{12}(A) = m_1 \oplus m_2 = \frac{1}{1 - c} \sum_{B \cap C = A} m_1(B)m_2(C) \tag{5}$$

where the conflict c can be computed as: $c = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$

In the DS, the independence of the available evidences is typically a strong constraint that can be relaxed by using further combinations rules, e.g. the *Dubois-Prade’s rule* [18].

$$m_{12}(A) = \sum_{B \cup C = A} m_1(B)m_2(C) \tag{6}$$

Differently from the Dempster’s rule, the latter considers the union between two sets of hypothesis rather than their intersection. As a result, the conflict between sources of information does not exists.

3 Evidence-Based Ensemble Learning for Description Logic

The TDT (and RF) learning approach is now recalled before introducing the method for the induction of an evidence-based versions of these classification models.

3.1 Class-Imbalance and Terminological Random Forests

In machine learning, the class-imbalance problem concerns the skewness of training data distribution. Considering a multilabel setting, where the number of class label is greater than 3, the problem usually occurs when the number of training instances belonging to the a particular class (the *majority class*) overwhelms the number of those belonging to the other classes (which represent the *majority class*). In order to tackle the problem, most common strategies based

on sampling strategy have been proposed [19]. One of the simplest method is an under-sampling strategy that randomly discards instances belonging to the majority class in order to re-balance the dataset. However, this method causes a loss of information due to the possible discarding of useful examples required for inducing a quite predictive model. A Terminological Random Forest (TRF) is an ensemble model trained through a procedure that combines a random under-sampling strategy with the ensemble learning induction [5]. The main purpose for the induction of these models is to mitigate the loss of information mentioned above in the context of SW knowledge bases. A TRF is basically made up of a certain number of Terminological Decision Trees (TDTs) [6], where each of them is built by considering a (quasi-)balanced dataset. The ensemble model assigns the final class for a new individual by appealing to a majority vote procedure. Therefore each TDT returns an *hard prediction*: this means that each tree contributes equally to the decision concerning the class label, regardless its confidence about predictions. In order to consider also this kind of information and tackling sundry problems as the uncertainty about the class assignment (i.e. when the confidence about either a class or another one is approximately equals) and the disagreement between classifiers that may lead to misclassifications [5], we need to resort to other models for the ensemble approach, such as Evidential Terminological Decision Trees [13].

3.2 Evidential Terminological Decision Trees

In [13], it has been shown how the class-membership prediction task can be tackled by inducing Evidential Terminological Decision Trees (ETDTs), an extension of the TDTs [6] based on evidential reasoning. ETDTs are defined in a similar way of TDTs. However, unlike TDTs, each node contains a couple $\langle D, m \rangle$, where D is a DL concept description and m is BBA concerning the membership w.r.t. D , rather than the sole concept description. Practically, to learn an ETDT model, a set of concept descriptions is generated from the current node by resorting to the refinement operator, denoted by ρ . For each concept, a BBA is also computed by considering the positive, negative and uncertain instances w.r.t. the generated concept. Then the best description (and the corresponding BBA) is selected, i.e. the one having the smallest non-specificity measure value w.r.t. the previous level. In other words, this means that the description is the one having the most definite membership.

Figure 1 reports a simple example of ETDT used for predicting whether a car is to be sent back to the factory (SendBack) or can be repaired. We can observe that the root concept $\exists \text{hasPart.T}$ is progressively specialized. Additionally, the concepts installed into the intermediate nodes are characterized by a decreasingly non specificity measure value.

3.3 Evidential Terminological Random Forests

An Evidential Terminological Random Forest (ETRF) is an ensemble of ETDTs. We will focus on the procedures for producing an ETRF and for predicting

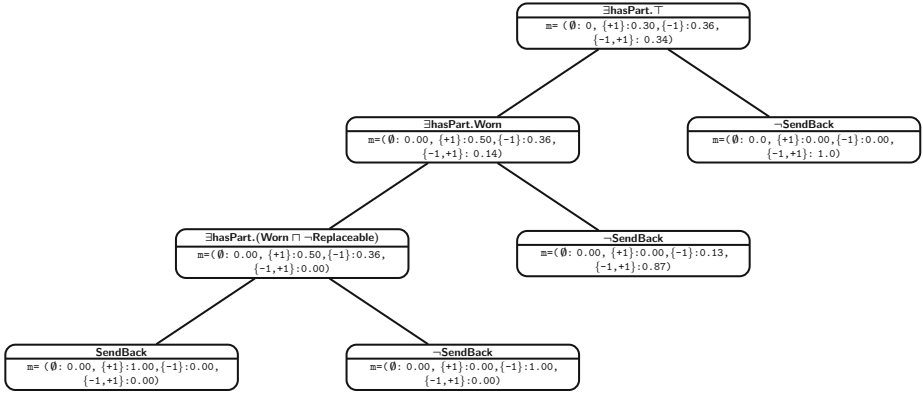


Fig. 1. A simple example of ETDT: each nodes contains a DL concept description and a BBA obtained by counting the instances that reach the node during the training phase

class-membership of input individuals exploiting an ETRF. Moving from the formulation of the concept learning problem proposed in [5], we will use the label set $\mathcal{L} = \{-1, +1\}$ as *frame of discernment* of the problem. The labels in \mathcal{L} are usually used to denote, respectively, the cases of positive and negative membership w.r.t. a target concept C . However, in order to represent the uncertain-membership related to the Open World Assumption, we will employ the label set $\mathcal{L}' = 2^{\mathcal{L}} \setminus \{\emptyset\}$ and the singletons $\{+1\}$ and $\{-1\}$ to denote the positive and negative membership w.r.t. C while the case of uncertain-membership will be labeled by $\mathcal{L} = \{-1, +1\}$.

Growing ETRFs. Algorithm 1 describes the procedure for producing an ETRF. In order to do this, the target concept C , a training set $\text{Tr} \subseteq \text{Ind}(\mathcal{A})$ and the desired number of trees n are required. Tr may contain not only positive and negative examples but also instances with uncertain membership w.r.t. C . According to a bagging approach, the training individuals are sampled with replacement in order to obtain n subsets $D_i \subseteq \text{Tr}$, with $i = 1, \dots, n$. In order to obtain D_i s, it is possible to apply various sampling strategies although, in this work, we followed the approach proposed in [5]. Firstly, the initial data distribution is considered by adopting a stratified sampling w.r.t. the class-membership values in order to represent instances of the minority class. In the second phase, undersampling can be performed on the training set in order to obtain (quasi-)balanced D_i sets (i.e. with a class imbalance that will not affect much the training process). This means that if the majority class is the negative one, the exceeding part of the counterexamples is randomly discarded. In the dual case, positive instances are removed. In addition, the sampling procedure removes also all the uncertain instances. In Algorithm 1, the procedure that returns the sets D_i implementing this strategy is `BALANCEDBOOTSTRAPSAMPLE`. For each D_i , an ETDT T is built by means of a recursive strategy, as described in [13] which is implemented by the procedure

Algorithm 1. The routines for inducing an ETRF

```

1  const:  $\theta$ : threshold
2  function INDUCEETRF( $\text{Tr}$  : training set;  $C$  : concept;  $n \in \mathbb{N}$ ): TRF
3  begin
4   $\hat{P}r \leftarrow \text{ESTIMATEPRIORS}(\text{Tr}, C)$ :  $\{C$  prior membership probability estimates $\}$ 
5   $F \leftarrow \emptyset$ 
6  for  $i \leftarrow 1$  to  $n$ 
7     $D_i \leftarrow \text{BALANCEDBOOTSTRAPSAMPLE}(\text{Tr})$ 
8    let  $D_i = \langle Ps, Ns, Us \rangle$ 
9     $T_i \leftarrow \text{INDUCEETDTREE}(D_i, C, \hat{P}r)$ ;
10    $F \leftarrow F \cup \{T_i\}$ 
11 return  $F$ 
12 end
13
14 function INDUCEETDTREE( $\langle Ps, Ns, Us \rangle$ : training set;  $C$ :concept;  $m$ : BBA,  $\hat{P}r$ : priors)
15 begin
16
17  $T \leftarrow \text{new ETDT}$ 
18 if  $|Ps| = 0$  and  $|Ns| = 0$  then
19   begin
20     if  $P_r(+1) \geq P_r(-1)$  then {pre-defined constants wrt the whole training set}
21        $T.\text{root} \leftarrow \langle C, m \rangle$ 
22     else
23        $T.\text{root} \leftarrow \langle \neg C, m \rangle$ 
24     return  $T$ 
25   end
26 if  $(m(\{-1\}) \simeq 0)$  and  $(m(\{+1\}) > \theta)$  then
27   begin
28      $T.\text{root} \leftarrow \langle C, m \rangle$ 
29   return  $T$ 
30   end
31 if  $(m(\{+1\}) \simeq 0)$  and  $(m(\{-1\}) > \theta)$  then
32   begin
33      $T.\text{root} \leftarrow \langle \neg C, m \rangle$ 
34   return  $T$ 
35   end
36  $RS \leftarrow \text{RANDOMSELECTION}(\rho(D))$  {random selection of specializations}
37  $S \leftarrow \emptyset$ 
38 for  $E \in RS$  {assignBBA for each candidate}
39    $m' \leftarrow \text{COMPUTE BBA}(E, \langle Ps, Ns, Us \rangle)$ 
40    $S \leftarrow S \cup \{E, m'\}$ 
41
42  $\langle E^*, m^* \rangle \leftarrow \text{SELECTBESTCANDIDATE}(S)$ 
43  $\langle \langle P^l, N^l, U^l \rangle, \langle P^r, N^r, U^r \rangle \rangle \leftarrow \text{SPLIT}(E^*, \langle Ps, Ns, Us \rangle)$ 
44  $T.\text{root} \leftarrow \langle E^*, m^* \rangle$ 
45  $T.\text{left} \leftarrow \text{INDUCEETDT}(\langle P^l, N^l, U^l \rangle, E^*, \hat{P}r)$ 
46  $T.\text{right} \leftarrow \text{INDUCEETDT}(\langle P^r, N^r, U^r \rangle, E^*, \hat{P}r)$ 
47 return  $T$ 
48 end

```

INDUCEETDT). It distinguishes various cases. The first one uses prior probability (estimate) to cope with the lack of examples ($|Ps| = 0$ and $|Ns| = 0$). The second one sets the class label for a leaf node if it is sufficiently pure, i.e. no positive (resp. negative) example is found while most examples are negative (resp. positive). This purity condition is evaluated by considering the BBA m given as input for the algorithm ($m(\{-1\}) \simeq 0$ and $m(\{+1\}) > \theta$, $m(\{+1\}) \simeq 0$ and $m(\{-1\}) > \theta$). The values of a BBA function for the membership values are obtained by computing the number of positive, negative and uncertain-membership instances w.r.t. the current concept. Finally, the third (recursive) case concerns the availability of both

Algorithm 2. Class-membership prediction

```

1  function CLASSIFYBYTRF( $a$  : individual;  $F$  : TRF;  $C$  : target concept) :  $\mathcal{L}$ 
2  begin
3   $M[] \leftarrow$  new array
4  for each  $T \in F$ 
5   $M[T] \leftarrow$  CLASSIFY( $a, T$ )
6
7   $\bar{m} \leftarrow \bigoplus_{m \in M} m$  {pooling according to a combination rule}
8
9  for each  $l \in 2^{\mathcal{L}}$  {class assignement}
10 Compute  $\overline{Bel}(l)$  from  $\bar{m}$ 
11
12 if ( $|\overline{Bel}(\{-1\}) - \overline{Bel}(\{+1\})| > \epsilon$ ) then
13 return  $\arg \max_{l \in \mathcal{L}' \setminus \{-1, +1\}} \overline{Bel}(l)$ 
14 else
15 return  $\mathcal{L}$ 
16 end
17
18 function CLASSIFY( $a, T$ ):  $\bar{m}$ 
19 begin
20  $L \leftarrow$  FINDLEAVES( $a, T$ ) {list of BBA}
21  $\bar{m} \leftarrow \bigoplus_{m \in L} m$ 
22 return  $\bar{m}$ 
23 end

```

negative and positive examples. In this case, the current concept description D has to be specialized by means of an operator exploring the search space of downward refinements of D . Following the approach described in [5, 8], the refinement step produces a set of candidate specializations $\rho(D)$ and a subset of them, namely RS , is then randomly selected (via function RANDOMSELECTION) by setting its cardinality according to the value returned by a function f applied to the cardinality of the set of specializations returned by the refinement operator (e.g. $\sqrt{|\rho(D)|}$). A BBA m' is then built for each candidate $E \in RS$. Again, the function can be obtained by counting the number of positive, negative and uncertain-membership instances). Then the best pair $\langle E^*, m^* \rangle \in S$ according to the non-specificity measure employed in [13] is determined by the SELECTBESTCANDIDATE procedure and finally installed in the current node. Specifically, the procedure tries to find the pair $\langle E^*, m^* \rangle$ having the smallest non-specificity measure value. After the assessment of the best pair E^* , the individuals are partitioned by the procedure SPLIT for the left or right branch according to the result of the instance-check w.r.t. E^* , maintaining the same group ($P^{l/r}, N^{l/r}$, or $U^{l/r}$). Note that a training example a is replicated in both children in case both $\mathcal{K} \not\models E^*(a)$ and $\mathcal{K} \not\models \neg E^*(a)$. The divide-and-conquer strategy is applied recursively until the instances routed to a node satisfy one of the stopping conditions discussed above.

Prediction. After an ETRF is produced, predictions can be made relying on the resulting classification model. The related procedure sketched in Algorithm 2 works as follows. Given the individual to be classified, for each tree T_i of the forest, the procedure CLASSIFY returns a BBA assigned to the leaves reached from the root in a path down the tree. Specifically, the algorithm traverses recursively

the ETDT by performing an instance check w.r.t. the concept contained in each node that is reached: let $a \in \text{Ind}(\mathcal{A})$ and D the concept installed in the current node, if $\mathcal{K} \models D(a)$ (resp. $\mathcal{K} \models \neg D(a)$) the left (resp. right) branch is followed. If neither $\mathcal{K} \models D(a)$ nor $\mathcal{K} \models \neg D(a)$ is verified, both branches are followed. After the exploration of a single ETDT, the list L may contain several BBAs. In this case, BBAs are pooled according to a combination rule as the Dubois-Prade's one [13]. The function CLASSIFY returns the combined BBA according to this rule (denoted by the symbol \oplus). After polling all trees, a set of BBAs deriving from the previous phase are exploited to decide the class label to the test individual a . Function CLASSIFYBYTRF takes an individual a and a forest F . Then, the algorithm iterates on the forest trees collecting the BBAs via function CLASSIFY. Then, the BBAs are pooled according to a further combination rule, which can be different from the one employed during the exploration of a single ETDT. Additionally, this combination rule should be also an associative operator [15]. In this way, the result should not be affected by the pooling order of the BBAs. In our experiments we combined these BBAs via Dempster's rules (denoted by the symbol \oplus in the function CLASSIFYBYTRF). By using this rule, the disagreement between classifiers, which corresponds to the conflict exploited as normalization factor, is explicitly considered by the meta-learner. The final decision is then made according to the belief function value computed from the pooled BBAs \bar{m} . In this case, we aim to select the $l \in 2^{\mathcal{L}}$ which maximizes the value of the function. However, in order to cope with the monotonicity of belief function which can lead easily to return an unknown-membership as a final prediction, the meta-learner must compare the value for the positive and negative class label and it assign the unknown membership if their values are approximately equal. This is made by comparing the difference between belief function values w.r.t. a threshold ϵ .

4 Preliminary Experiments

The experimental evaluation aims at evaluating the effectiveness of the classification based on the ETRF models¹ and the improvement in terms of prediction w.r.t. TRFs. We provide the details of the experimental setup and present and discuss the outcomes.

4.1 Setup

Various Web ontologies have been considered in the experiments (see Table 1). They are available on TONES repository². For each ontology of TONES, 15 query concepts have been randomly generated by combining (using the conjunction and disjunction operators or universal and existential restriction) 2 through 8 (primitive or defined) concepts of the ontology.

¹ The source code will be available at: <https://github.com/Giuseppe-Rizzo/SWML-Algorithms>.

² <http://www.inf.unibz.it/tones/index.php>.

Table 1. Ontologies employed in the experiments

Ontology	DL Lang.	#Concepts	#Roles	#Individuals
BCO	$\mathcal{ALCHO}\mathcal{F}(\mathcal{D})$	196	22	112
BioPAX	$\mathcal{ALC}\mathcal{L}\mathcal{I}\mathcal{F}(\mathcal{D})$	74	70	323
NTN	$\mathcal{SH}\mathcal{L}\mathcal{I}\mathcal{F}(\mathcal{D})$	47	27	676
HD	$\mathcal{ALC}\mathcal{L}\mathcal{I}\mathcal{F}(\mathcal{D})$	1498	10	639

As in previous works [5, 13], because of the limited population of the considered ontologies, all the individuals occurring in each ontology were employed as (training or test) examples.

A 10-fold cross validation design of the experiments was adopted so that the final results are averaged for each of the considered indices (see below). We compared our extensions with other tree-based classifiers: TDTs [6], TRFs [5] and ETDTs [13].

In order to learn each ETDTs by considering a balanced set of examples, a stratified sampling was required (see Sect. 3). Three stratified sampling rates related to the D_i s were set in our experiments, namely 50 %, 70 % and 80 %.

Finally, forests with an increasing number of trees were induced, namely: 10, 20 and 30. For each tree in a forest, the number of randomly selected candidates was determined as the square root of candidate refinements: $\sqrt{|\rho(\cdot)|}$. We employed these settings for training both ETRFs and TRFs. As in previous works [5, 6, 13], to compare the predictions made using RFs against the ground truth assessed by a reasoner, the following indices were computed:

- *match* rate (M%), i.e. test individuals for which the inductive model and a reasoner agree on the membership (both $\{+1\}$, $\{-1\}$, or $\{-1, +1\}$);
- *commission* rate (C%) i.e. test cases where the determined memberships are opposite (i.e. $\{+1\}$ vs. $\{-1\}$ or viceversa);
- *omission* rate (O%), i.e. test cases for which the inductive method cannot determine a definite membership while the reasoner can ($\{-1, +1\}$ vs. $\{+1\}$ or $\{-1\}$);
- *induction* rate (I%). i.e. test cases where the inductive method can predict a definite membership while the reasoner cannot assess it ($\{+1\}$ or $\{-1\}$ vs. $\{-1, +1\}$).

4.2 Results

As regards the distribution of the instances w.r.t. the target concepts, we observed that negative instances outnumber the positive ones in BCO and Human Disease (HD). In the case of BCO this occurred for all concepts but one with a ratio between positive and negative instances of 1 : 20. In the case of HD this kind of imbalance occurred for all the queries. Moreover, in the case of HD the number of instances with an uncertain-membership is very large (about 90 %). On the other hand, in the case of NTN, we noted the predominance of positive instances: for

Table 2. Results of experiments with TDTs and ETDTs models

<i>Ontology index</i>		TDT	ETDTs
BCO	M%	80.44 ± 11.01	90.31 ± 14.79
	C%	07.56 ± 08.08	01.86 ± 02.61
	O%	05.04 ± 04.28	00.00 ± 00.00
	I%	06.96 ± 05.97	07.83 ± 15.35
BIOPAX	M%	66.63 ± 14.60	87.00 ± 07.15
	C%	31.03 ± 12.95	11.57 ± 02.62
	O%	00.39 ± 00.61	00.00 ± 00.00
	I%	01.95 ± 07.13	01.43 ± 08.32
NTN	M%	68.85 ± 13.23	23.87 ± 26.18
	C%	00.37 ± 00.30	00.00 ± 00.00
	O%	09.51 ± 07.06	00.00 ± 00.00
	I%	21.27 ± 08.73	75.13 ± 26.18
HD	M%	58.31 ± 14.06	10.69 ± 01.47
	C%	00.44 ± 00.47	00.07 ± 00.17
	O%	05.51 ± 01.81	00.00 ± 00.00
	I%	35.74 ± 15.90	89.24 ± 01.46

Table 3. Comparison between TRFs and ETRF with sampling rate of 50 %

<i>Ontology index</i>		<i>Sampling rate 50 %</i>					
		TRF			ETRF		
		<i>10 trees</i>	<i>20 trees</i>	<i>30 trees</i>	<i>10 trees</i>	<i>20 trees</i>	<i>30 trees</i>
BCO	M%	86.27 ± 15.79	86.24 ± 15.94	86.26 ± 15.84	91.31 ± 06.35	91.31 ± 06.35	91.31 ± 06.35
	C%	02.47 ± 03.70	02.43 ± 03.70	02.84 ± 03.70	02.91 ± 02.45	02.91 ± 02.45	02.91 ± 02.45
	O%	01.90 ± 07.30	01.97 ± 07.55	01.92 ± 07.37	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	09.36 ± 13.96	09.36 ± 13.96	09.36 ± 13.96	05.88 ± 06.49	05.88 ± 06.49	05.88 ± 06.49
BIOPAX	M%	75.30 ± 16.23	75.30 ± 16.23	75.30 ± 16.23	96.92 ± 08.07	96.79 ± 08.15	96.55 ± 08.15
	C%	18.74 ± 17.80	18.74 ± 17.80	18.74 ± 17.80	00.79 ± 01.22	00.91 ± 01.74	00.77 ± 01.74
	O%	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	01.97 ± 07.16	01.97 ± 07.16	01.97 ± 07.16	02.29 ± 08.13	02.30 ± 08.15	02.30 ± 08.15
NTN	M%	83.41 ± 07.85	83.42 ± 07.85	83.42 ± 07.85	05.38 ± 07.38	05.38 ± 07.38	05.38 ± 07.38
	C%	00.02 ± 00.04	00.02 ± 00.04	00.02 ± 00.04	06.58 ± 07.51	06.58 ± 07.51	06.58 ± 07.51
	O%	13.40 ± 10.17	13.40 ± 10.17	13.40 ± 10.17	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	03.17 ± 04.65	03.16 ± 04.65	03.16 ± 04.65	88.05 ± 08.50	88.05 ± 08.50	88.05 ± 08.50
HD	M%	68.00 ± 16.98	68.00 ± 16.99	67.98 ± 16.99	10.29 ± 00.00	10.29 ± 00.01	10.29 ± 00.02
	C%	00.02 ± 00.05	00.02 ± 00.05	00.02 ± 00.05	00.26 ± 00.26	00.26 ± 00.27	00.26 ± 00.28
	O%	06.38 ± 02.03	06.38 ± 02.03	06.38 ± 02.03	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	25.59 ± 18.98	25.59 ± 18.98	25.59 ± 18.98	89.24 ± 00.26	89.24 ± 00.26	89.24 ± 00.26

most concepts the ratio between positive and negative instances was 12 : 1 and a lot of uncertain-membership instances were found (again, over 90 %). A weaker imbalance could be noted with BIOPAX. For most query concepts the ratio between positive and negative instances was 1 : 5. In addition, for most query concepts, uncertain-membership instances lacked. This kind of instances were available only for 2 queries. The class distribution was balanced for three concepts only.

Tables 2, 3, 4 and 5 report the results of this empirical evaluation. On the other hand, Table 6 shows the differences between indexes for TRFs and ETRFs. In general, we can observe how ensemble methods perform better or, in the worst cases, have the same performance of a single classifiers approach for most ontologies. For example, when we compare ETRFs w.r.t. ETDTs, a significant improvement was obtained for BIOPAX (the match rate was around 96 % for ETRFs and 87 % for ETDTs). For BCO, there was a more limited improvement:

Table 4. Comparison between TRFs and ETRF with sampling rate of 70 %

Ontology index		Sampling rate 70 %					
		TRF			ETRF		
		10 trees	20 trees	30 trees	10 trees	20 trees	30 trees
BCO	M%	84.12 ± 18.27	85.70 ± 16.98	85.52 ± 17.09	91.31 ± 06.35	91.31 ± 06.35	91.31 ± 06.35
	C%	02.16 ± 03.09	02.32 ± 03.39	02.30 ± 03.38	02.91 ± 02.45	02.91 ± 02.45	02.91 ± 02.45
	O%	04.50 ± 12.59	02.65 ± 09.93	02.86 ± 10.04	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	09.23 ± 13.99	09.33 ± 13.97	09.31 ± 13.91	05.88 ± 06.49	05.88 ± 06.49	05.88 ± 06.49
BIOPAX	M%	75.30 ± 16.23	75.30 ± 16.23	75.30 ± 16.23	96.65 ± 08.05	95.98 ± 08.13	96.55 ± 08.15
	C%	18.74 ± 17.80	18.74 ± 17.80	18.74 ± 17.80	01.07 ± 01.67	01.71 ± 02.50	00.77 ± 01.74
	O%	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	01.97 ± 07.16	01.97 ± 07.16	01.97 ± 07.16	02.28 ± 08.13	02.31 ± 08.17	02.30 ± 08.15
NTN	M%	83.42 ± 07.85	83.42 ± 07.85	83.42 ± 07.85	05.50 ± 07.28	05.50 ± 07.28	05.50 ± 07.28
	C%	00.02 ± 00.04	00.02 ± 00.04	00.02 ± 00.04	06.52 ± 07.54	06.52 ± 07.54	06.52 ± 07.54
	O%	13.40 ± 10.17	13.40 ± 10.17	13.40 ± 10.17	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	03.16 ± 04.65	03.16 ± 04.65	03.16 ± 04.65	87.99 ± 08.84	87.99 ± 08.84	87.99 ± 08.84
HD	M%	68.00 ± 16.98	68.00 ± 16.99	67.98 ± 16.99	10.29 ± 00.00	10.29 ± 00.01	10.29 ± 00.02
	C%	00.02 ± 00.05	00.02 ± 00.05	00.02 ± 00.05	00.26 ± 00.26	00.26 ± 00.27	00.26 ± 00.28
	O%	06.38 ± 02.03	06.38 ± 02.03	06.38 ± 02.03	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	25.59 ± 18.98	25.59 ± 18.98	25.59 ± 18.98	89.24 ± 00.26	89.24 ± 00.26	89.24 ± 00.26

Table 5. Comparison between TRFs and ETRF with sampling rate of 80 %

Ontology index		Sampling rate 80 %					
		TRF			ETRF		
		10 trees	20 trees	30 trees	10 trees	20 trees	30 trees
BCO	M%	75.57 ± 24.28	81.27 ± 19.27	79.33 ± 22.41	91.31 ± 06.35	91.31 ± 06.35	91.31 ± 06.35
	C%	01.45 ± 01.77	01.89 ± 02.65	01.64 ± 02.36	02.91 ± 02.45	02.91 ± 02.45	02.91 ± 02.45
	O%	13.51 ± 22.19	08.05 ± 15.04	10.38 ± 19.28	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	08.47 ± 14.07	08.65 ± 14.23	09.36 ± 13.96	05.88 ± 06.49	05.88 ± 06.49	05.88 ± 06.49
BIOPAX	M%	75.30 ± 16.23	75.30 ± 16.23	75.30 ± 16.23	95.47 ± 07.95	94.29 ± 08.96	96.55 ± 08.15
	C%	18.74 ± 17.80	18.74 ± 17.80	18.74 ± 17.80	02.24 ± 02.63	03.40 ± 05.54	00.77 ± 01.74
	O%	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	01.97 ± 07.16	01.97 ± 07.16	01.97 ± 07.16	02.29 ± 08.11	02.31 ± 08.17	02.30 ± 08.15
NTN	M%	83.41 ± 07.85	83.42 ± 07.85	83.42 ± 07.85	05.50 ± 07.28	05.50 ± 07.28	05.50 ± 07.28
	C%	00.02 ± 00.04	00.02 ± 00.04	00.02 ± 00.04	06.52 ± 07.54	06.52 ± 07.54	06.52 ± 07.54
	O%	13.40 ± 10.17	13.40 ± 10.17	13.40 ± 10.17	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	03.17 ± 04.65	03.16 ± 04.65	03.16 ± 04.65	87.99 ± 08.84	87.99 ± 08.84	87.99 ± 08.84
HD	M%	68.00 ± 16.98	68.00 ± 16.99	67.98 ± 16.99	10.29 ± 00.00	10.29 ± 00.01	10.29 ± 00.02
	C%	00.02 ± 00.05	00.02 ± 00.05	00.02 ± 00.05	00.26 ± 00.26	00.26 ± 00.27	00.26 ± 00.28
	O%	06.38 ± 02.03	06.38 ± 02.03	06.38 ± 02.03	00.00 ± 00.00	00.00 ± 00.00	00.00 ± 00.00
	I%	25.59 ± 18.98	25.59 ± 18.98	25.59 ± 18.98	89.24 ± 00.26	89.24 ± 00.26	89.24 ± 00.26

it was only around 1.31 % and it was likely due to the number of examples available in BCO. In this case, when ETRFs model were induced, there was a larger overlap between the ETDTs in the forests and the sole ETDT model employed in the single-classifier approach, i.e. the models were very similar to each other.

As regards the comparison between ETRFs and TRFs model, an improvement of match rate and a subsequent decrease of induction rate was observed for BCO. This improvement was around 6 % for match rate while it was of 3 % for the induction rate when a sampling rate of 50 % was employed. The improvement of match rate was larger when the sampling rate of 70 % and 80 % were employed. In this case, the addition of further instances lead to that the improvement of the predictiveness of the ETRFs. The ensemble of models proposed in this

Table 6. Differences between the results for TRFs and ETRFs model. The symbol ● is used to denote that a positive or negative difference that is in favor of ETRFs, while the symbol ○ is used to denote a positive or negative difference that is in favor of TRFs

Ontology	Index	Sampling rate 50 %			Sampling rate 70 %			Sampling rate 80 %		
		10 trees	20 trees	30 trees	10 trees	20 trees	30 trees	10 trees	20 trees	30 trees
BCO	$\Delta M\%$	+05.04 ●	+05.07 ●	+05.05 ●	+07.19 ●	+05.61 ●	+05.79 ●	+15.74 ●	+10.04 ●	+11.98 ●
	$\Delta C\%$	+00.44 ○	+00.48 ○	+00.07 ○	+00.75 ○	+00.59 ○	+00.61 ○	+01.46 ○	+01.02 ○	+01.27 ○
	$\Delta O\%$	-01.90 ●	-01.97 ●	-01.92 ●	-04.50 ●	-02.65 ●	-02.86 ●	-13.51 ●	-8.05 ●	-10.38 ●
	$\Delta I\%$	-03.48 ●	-03.48 ●	-03.48 ●	-03.35 ●	-03.45 ●	-03.43 ●	-02.59 ●	-02.77 ●	-03.48 ●
BIO-PAX	$\Delta M\%$	+21.62 ●	+21.49 ●	+21.25 ●	+21.35 ●	+20.68 ●	+21.25 ●	+20.17 ●	+18.99 ●	+21.25 ●
	$\Delta C\%$	-17.95 ●	-17.83 ●	-17.97 ●	-17.67 ●	-17.03 ●	-17.97 ●	-16.50 ●	-15.34 ●	-17.97 ●
	$\Delta O\%$	+00.00 ●	+00.00 ●	+00.00 ●	+00.00 ●	+00.00 ●	+00.00 ●	+00.00 ●	+00.00 ●	+00.00 ●
	$\Delta I\%$	+00.32 ○	+00.33 ○	+00.33 ○	+00.31 ○	+00.34 ○	+00.33 ○	+00.32 ○	+00.34 ○	+00.33 ○
NTN	$\Delta M\%$	-78.03 ○	-78.04 ○	-78.04 ○	-77.92 ○	-77.92 ○	-77.92 ○	-77.91 ○	-77.92 ○	-77.92 ○
	$\Delta C\%$	+06.56 ○	+06.56 ○	+06.56 ○	+06.50 ○	+06.50 ○	+06.50 ○	+06.50 ○	+06.50 ○	+06.50 ○
	$\Delta O\%$	-13.40 ●	-13.40 ●	-13.40 ●	-13.40 ●	-13.40 ●	-13.40 ●	-13.40 ●	-13.40 ●	-13.40 ●
	$\Delta I\%$	+84.88 ○	+84.89 ○	+84.89 ○	+84.83 ○	+84.83 ○	+84.83 ○	+84.82 ○	+84.83 ○	+84.83 ○
HD	$\Delta M\%$	-57.71 ○	-57.71 ○	-57.69 ○	-57.71 ○	-57.71 ○	-57.69 ○	-57.71 ○	-57.71 ○	-57.69 ○
	$\Delta C\%$	+00.24 ○	+00.24 ○	+00.24 ○	+00.24 ○	+00.24 ○	+00.24 ○	+00.24 ○	+00.24 ○	+00.24 ○
	$\Delta O\%$	-06.38 ●	-06.38 ●	-06.38 ●	-06.38 ●	-06.38 ●	-06.38 ●	-06.38 ●	-06.38 ●	-06.38 ●
	$\Delta I\%$	+63.65 ○	+63.65 ○	+63.65 ○	+63.65 ○	+63.65 ○	+63.65 ○	+63.65 ○	+63.65 ○	+63.65 ○

paper showed a more conservative behavior w.r.t. the original version. It can be noted that the increase of match rate was mainly due to uncertain-membership instances that were not classified as induction cases, as a result of the values of belief functions employed for making decisions. Another cause is related to the lack of omission cases. In this case, the procedure for forcing the answer leads to decide in favor of the correct class-membership value. Besides the value of commission rate did not change in a significant way. The proposed extension is also more stable in terms of standard deviation: for ETRFs, this value is lower than the one obtained for TRFs.

With BIO-PAX, we observed again the increase of the match and a significant decrease of commission rate. Also the induction rate was larger with ETRFs than with TRFs, likely due to the procedure for forcing the answer. As regards the experiments on HD and NTN ontology, we can observe, differently from the original version of TRFs, how the induction rate was very high when ETRFs were employed. For the latter case, this result was mainly due to the original data distribution that showed an overwhelming of uncertain instances. As previously mentioned, they approximately represented about 50 % of the total number of instances in the ABox of HD and about 90 % for NTN. TRFs showed a conservative behavior by returning an unknown membership (due to uncertain results of the intermediate tests during the exploration of trees [5]) which tends to preserve the matches with the gold-standard membership also in case of uncertain membership. This explains the high match rate observed in the experiments. After the induction of ETRFs, the models showed a braver behavior also due to the forcing procedure. As a result, it tends to more easily assign a positive or negative membership to a test instance leading to the increase of the induction rate, with a value of about 89 % while omission cases missed. Induction cases represent new non-derivable knowledge that can be potentially useful for ontology completion, their larger number suggest that the

result may be also due to the existing noise (also due to the employment of the entire ABox as dataset). This basically means that most induced assertions may be not definitely related to learned concepts, but they cannot be considered as real errors like commission rate.

Similarly to our previous experiments proposed in [5], we observed also how the generated concept descriptions that were installed as node for each ETDT do not improve the quality of the splittings, similarly to the case of TDTs where the training was lead by the information gain criterion. This occurred for all the datasets that were considered here. In both cases, most instances were sent along a branch, while a small number of them were sent along the other one. This means that *small disjuncts problem* is a common problem both TRFs and ETRFs and neither the information gain nor the non-specificity measure can be considered as suitable measures for selecting the best concept description that is used to split instances during the training phase. A further remark concerns the predictiveness of the proposed method w.r.t. both the sampling methods and the number of trees in a forest. Also for ETRFs, the performance did not change significantly when a larger number of trees was set or when the algorithm resort to a larger stratified sampling rate. While in the former case the results are likely due to a weak diversification between ETDTs, in the latter case, the result was likely due to the availability of examples whose employment did not change the quality of splittings generated during the growth process. For ETRFs, similarly to TRF models, the refinement operator is still a bottleneck for learning phase: execution times spanned from few minutes to almost 10h as the experiments proposed in [5]. However, when an intermediate test with an uncertain result was encountered, the exploration of alternative paths affected the efficiency of the proposed method.

5 Conclusion and Extensions

We have proposed an algorithm for inducing Evidential Terminological Random Forests, an extension of Terminological Random Forests devised to tackle the class-imbalance problem for learning predictive classification models for SW knowledge bases. As the original version, the algorithm combines a sampling approach with ensemble learning techniques. The resulting models combine predictions that are represented as basic belief functions rather than votes by exploiting combination rules in the context of the Dempster-Shafer Theory for making the final decision. In addition, a preliminary empirical evaluation with publicly available ontologies has been performed. The experiments have shown how the new classification model seems to be more predictive than the previous ones and it tends to assign a definite membership. Besides, the predictiveness of the model can be sufficiently tolerant to variation of the number of trees and the sampling rate. The standard deviation is also lower than the original TRFs. In the future, we plan to extend the method along various directions. One regards the choice of the refinement operator that may be applied in order to generate more discriminative intermediate tests. This plays a crucial role for the quality of the classifiers involved in the ensemble model

in order to obtain quite predictive weak learners from both expressive and shallow ontologies extracted from the Linked Data cloud [20]. In order to cope with the latter case, the method could be parallelized in order to employ it as a non-standard tool to reason over such datasets. Further ensemble techniques and novel rules for combining the answers of the weak learners could be employed. For example, weak learners can be induced from subsets of training instances generated by means of a procedure based on cross-validation rather than sampling with replacement. Finally, further investigations may concern the application of strategies aiming to optimize the ensemble, that is an important characteristic of such learning methods [12, 21].

Acknowledgments. This work fulfills the objectives of the LOGIN “LOGistica INTe-grata” (2012–2015) (PII INDUSTRY 2015), announcement “New Technologies for the Made in Italy”.

References

1. Rettinger, A., Lösch, U., Tresp, V., d’Amato, C., Fanizzi, N.: Mining the semantic web - statistical learning for next generation knowledge bases. *Data Min. Knowl. Discov.* **24**, 613–662 (2012)
2. d’Amato, C., Fanizzi, N., Esposito, F.: Inductive learning for the semantic web: what does it buy? *Semant. Web* **1**, 53–59 (2010)
3. d’Amato, C., Fanizzi, N., Fazzinga, B., Gottlob, G., Lukasiewicz, T.: Ontology-based semantic search on the web and its combination with the power of inductive reasoning. *Ann. Math. Artif. Intell.* **65**, 83–121 (2012)
4. He, H., Ma, Y.: *Imbalanced Learning: Foundations, Algorithms, and Applications*, 1st edn. Wiley-IEEE Press, Hoboken (2013)
5. Rizzo, G., d’Amato, C., Fanizzi, N., Esposito, F.: Tackling the class-imbalance learning problem in semantic web knowledge bases. In: Janowicz, K., Schlobach, S., Lambrix, P., Hyvönen, E. (eds.) *EKAW 2014. LNCS*, vol. 8876, pp. 453–468. Springer, Heidelberg (2014)
6. Fanizzi, N., d’Amato, C., Esposito, F.: Induction of concepts in web ontologies through terminological decision trees. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) *ECML PKDD 2010, Part I. LNCS*, vol. 6321, pp. 442–457. Springer, Heidelberg (2010)
7. Breiman, L.: Random forests. *Mach. Learn.* **45**, 5–32 (2001)
8. Assche, A.V., Vens, C., Blockeel, H., Dzeroski, S.: First order random forests: learning relational classifiers with complex aggregates. *Mach. Learn.* **64**, 149–182 (2006)
9. Kuncheva, L.: A theoretical study on six classifier fusion strategies. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 281–286 (2002)
10. Xu, L., Krzyzak, A., Suen, C.: Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. Syst. Man Cybern.* **22**, 418–435 (1992)
11. Rogova, G.: Combining the results of several neural network classifiers. In: Yager, R., Liu, L. (eds.) *Classic Works of the Dempster-Shafer Theory of Belief Functions. Studies in Fuzziness and Soft Computing*, vol. 219, pp. 683–692. Springer, Heidelberg (2008)

12. Yin, X.C., Yang, C., Hao, H.W.: Learning to diversify via weighted kernels for classifier ensemble. CoRR abs/1406.1167 (2014)
13. Rizzo, G., d'Amato, C., Fanizzi, N., Esposito, F.: Towards evidence-based terminological decision trees. In: Laurent, A., Strauss, O., Bouchon-Meunier, B., Yager, R.R. (eds.) IPMU 2014, Part I. CCIS, vol. 442, pp. 36–45. Springer, Heidelberg (2014)
14. Bi, Y., Guan, J., Bell, D.: The combination of multiple classifiers using an evidential reasoning approach. *Artif. Intell.* **172**, 1731–1751 (2008)
15. Sentz, K., Ferson, S.: Combination of evidence in Dempster-Shafer theory. Technical report, SANDIA, SAND2002-0835 (2002)
16. Klir, J.: *Uncertainty and Information*. Wiley, Hoboken (2006)
17. Smarandache, F., Dezert, J.: An introduction to the DS_m theory for the combination of paradoxical, uncertain, and imprecise sources of information. CoRR abs/cs/0608002 (2006)
18. Dubois, D., Prade, H.: On the combination of evidence in various mathematical frameworks. In: Flamm, J., Luisi, T. (eds.) *Reliability Data Collection and Analysis*. Eurocourses, vol. 3, pp. 213–241. Springer, The Netherlands (1992)
19. He, H., Garcia, E.A.: Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**, 1263–1284 (2009)
20. Heath, T., Bizer, C.: *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web. Morgan and Claypool Publishers, San Rafael (2011)
21. Fu, B., Wang, Z., Pan, R., Xu, G., Dolog, P.: An integrated pruning criterion for ensemble learning based on classification accuracy and diversity. In: Uden, L., Herrera, F., Bajo, J., Corchado, J.M. (eds.) *7th International Conference on KMO. AISC*, vol. 172, pp. 47–58. Springer, Heidelberg (2013)