Inductive Classification through Evidence-based Models and Their Ensembles

Giuseppe Rizzo, Claudia d'Amato, Nicola Fanizzi and Floriana Esposito

Dipartimento di Informatica
Università degli Studi di Bari "Aldo Moro", Bari, Italy

ESWC 2015
June 3rd, 2015

Outline

Motivations

AIM: predicting the membership of an individual w.r.t. a query concept

- typically based on automated reasoning techniques
- Inferences are affected by the incompleteness of the Semantic Web
- decided using models induced by Machine learning methods
- The quality depends on the training data distribution
  - Given a query concept, generally, many uncertain-membership examples than individuals with a definite membership
  - **We are assuming a ternary classification problem**

Motivations

Previous solutions and the current limits

- We started to investigate the imbalance learning problem by resorting to a solution which combines (under-)sampling methods and ensemble learning models
  - for overcoming the loss of information due to the discarded instances
  - Terminological Decision Tree (TDT): a DL-based Decision Tree for concept learning and assertion prediction problems
  - combined to obtain Terminological Random Forests (TRF)
- Some limits:
  - predictions made according to simple majority vote procedure (no conflicts, no uncertainty are considered)
  - misclassifications mainly due to evenly distributed votes
- Further rules (meta-learner) required

Introduction & Motivations

Underlying idea

Using soft predictions (predictions with a confidence measure for each class value) obtained by each tree for weighting the votes

- TDTs return only hard predictions (i,.e. predicted class without any information)
- Dempster-Shafer Theory (DS) operators for information fusion
  - Solution: Resort and modify the Evidential TDTs (ETDTs)

The Dempster-Shafer Theory (DS)

- Frame of discernement Ω
  - a set of hypotheses for a domain, e.g. the membership values for an individual given a concept $\Omega = \{-1, +1\}$
- Basic Belief Assignement (BBA) $m : 2^{\Omega} \rightarrow [0, 1]$
  - the amount of belief exactly committed to $A \subseteq \Omega$
- Belief function: $\forall A, B \in 2^{\Omega} \quad Bel(A) = \sum_{B \subseteq A} m(B)$
- Plausibility function: $\forall A, B \in 2^{\Omega} \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$

The Dempster-Shafer Theory (DS)

- Combination rules: used for pooling evidences for the same frame of discernment coming from various sources of information
  - Dempster's rule
    - $\forall A, B, C \subseteq \Omega \quad m_{12}(A) = m_1 \oplus m_2 = \frac{1}{1-c} \sum_{B \cap C = A} m_1(B) m_2(C)$
  - Dubois-Prade's rule
    - $\forall A, B, C \subseteq \Omega \quad m_{12}(A) = \sum_{B \cup C = A} m_1(B) m_2(C)$

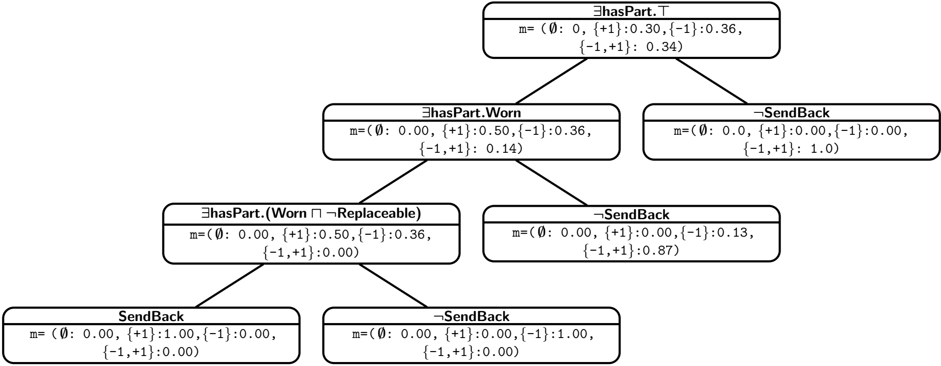
Evidential TDTs

An ETDT is a binary tree where:

- each node contains a conjunctive concept description D and a BBA m obtained by counting the positive, negative and uncertain instances;
- each departing edge is the result of instance-check test w.r.t. D, i.e., given an individual a, $\mathcal{K} \models D(a)$?
- a child node with the concept description D is obtained using a refinement operator

The model can be used for returning soft prediction

An example of ETDT



$\Omega = \{-1, +1\}$
$\{+1\} \leftrightarrow \mathcal{K} \models D(a)\ \forall a \in \mathrm{Ind}(A)$
$\{-1\} \leftrightarrow \mathcal{K} \models \neg D(b)\ \forall b \in \mathrm{Ind}(A)$
$\{-1, +1\}$ otherwise

Evidential Terminological Random Forests

- In order to tackle the imbalance learning problem, we propose Evidential Terminological Random Forest (ETRF), where
  - each ETDT returns a soft prediction in the form of BBA
  - the meta-learner is a combination rule

Learning Evidential Terminological Random Forests

Given:

- a target concept C
- the number of trees n
- a training set $\mathrm{Tr} = \langle \mathrm{Ps}, \mathrm{Ns}, \mathrm{Us} \rangle$
  - $Ps = \{a \in \mathrm{Ind}(\mathcal{A}) | \mathcal{K} \models C(a)\}$
  - $Ns = \{b \in \mathrm{Ind}(\mathcal{A}) | \mathcal{K} \models \neg C(b)\}$
  - $Us = \{c \in \mathrm{Ind}(\mathcal{A}) | \mathcal{K} \not\models C(c) \wedge \mathcal{K} \not\models \neg C(c)\}$

the algorithm can be summarized as follows:

- build a n bootstrap samples with a balanced distribution
- for each sample learn an ETDT model

Learning ETRF

Building bootstrap samples

1. a stratified sampling with replacement procedure is employed in order to represent the minority class instances in the bootstrap sample.
2. the majority class instances (either positive, negative and uncertain-membership instances) are discarded.

Learning ETRF

Learning ETDTs

- Divide-and-conquer algorithm for learning an ETDT [Rizzo et al.@IPMU, 2014]
- Steps:
  1. refinement of the concept description installed into the current node
  2. Random selection of a subset of candidates
  3. A BBA for each selected description
  4. The concept having the most definite membership (and its BBA) installed into the new node.
- Stop conditions: the node is pure w.r.t. the membership

Predicting membership for unseen individuals

- Given a forest F and a new individual a, the algorithm collects BBAs returned by each ETDT
- The BBA returned by an ETDT is decided by following a path according to the instance check test result.
- For a concept description installed as node D
  - if $\mathcal{K} \models D(a)$ the left branch is followed
  - if $\mathcal{K} \models \neg D(a)$ the right branch is followed
  - otherwise both branches are followed
  - Various leaves can be reached and the corresponding BBAs are pooled according to the combination rule

Predicting membership for unseen individuals

- The set of BBAs returning from all the ETDTs are combined through the combination rule
- After a pooled BBA $\overline{m}$ is obtained, Bel (resp. Pl) function is derived
- Final membership assignement: hypothesis which maximizes belief (resp. plausibility) function
  - Bel and Pl function are monotonic : uncertain-memberhip is more probable
  - Return the uncertain-membership value when the belief for the positive- and negative-membership are approximately equal

Experiments

- 15 query concepts randomly generated
- 10-fold cross validation
- number of candidates randomly selected: $\sqrt{|\rho(\cdot)|}$
- Comparison w.r.t. TDTs, ETDTs, TRFs
- Forest sizes: 10, 20, 30 trees
- Stratified Sampling rates: 50%, 70 %, 80 %
- Metrics:
  - match: individuals for which the inductive model and a reasoner predict the same membership
  - commission: cases of opposite predictions
  - omission: individuals having a definite membership that cannot be predicted inductively;
  - induction: predictions that are not logically derivable.

Some results...

Ontology index		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**

Some results...

Ontology	*index*	*Sampling rate 50 %* TRF *10 trees*	ETRF *10 trees*
BCO	M%	**86.27 ± 15.79**	**91.31 ± 06.35**
	C%	**02.47 ± 03.70**	**02.91 ± 02.45**
	O%	**01.90 ± 07.30**	**00.00 ± 00.00**
	I%	**09.36 ± 13.96**	**05.88 ± 06.49**
BIOPAX	M%	**75.30 ± 16.23**	**96.92 ± 08.07**
	C%	**18.74 ± 17.80**	**00.79 ± 01.22**
	O%	**00.00 ± 00.00**	**00.00 ± 00.00**
	I%	**01.97 ± 07.16**	**02.29 ± 08.13**
NTN	M%	**83.41 ± 07.85**	**05.38 ± 07.38**
	C%	**00.02 ± 00.04**	**06.58 ± 07.51**
	O%	**13.40 ± 10.17**	**00.00 ± 00.00**
	I%	**03.17 ± 04.65**	**88.05 ± 08.50**
HD	M%	**68.00 ± 16.98**	**10.29 ± 00.00**
	C%	**00.02 ± 00.05**	**00.26 ± 00.26**
	O%	**06.38 ± 02.03**	**00.00 ± 00.00**
	I%	**25.59 ± 18.98**	**89.24 ± 00.26**

Discussion

- improved performance of ETRFs w.r.t. the other models
  - higher match rate and induction rate
  - a lower standard deviation
- smallest changes of performance w.r.t. the forest size
  - weak diversification(overlapping) between trees by increasing the number of trees
- refinement operator is a bottleneck for learning phase

Conclusions and Further Extensions

- We proposed an ensemble solution based on DS to improve the predictiveness of the models for class-membership prediction with imbalanced training data distribution
- Extensions:
  - Development and reuse of refinement operators
  - Further ensemble techniques and combination rules
  - Experiments with larger ontologies
  - Parallelization of the current implementation

Thank you!

Questions?