Argumentation for Machine Learning

A Survey

Oana Cocarascu & Francesca Toni

September 16, 2016

Imperial College London

Introduction

Approaches using argumentation for ML

Supervised learning

- Argumentation-Based Machine Learning (ABML)
- Argument-Based Inductive Logic Programming (ABILP)
- Concept Learning as Argumentation (CLA)
- Multi-Agent Inductive Concept Learning (MAICL)
- Classification enhanced with Argumentation (CleAr)

Unsupervised learning

- Argumentation for ART (A-ART)

Reinforcement learning

- Argumentation Accelerated Reinforcement Learning (AARL)

Differences in approaches

- use of argumentation
- argumentation framework
- semantics
- single or multi-agent
- during or after ML
- different outcomes
  - improving performance
  - improving ML explanatory power
  - rendering the ML process more transparent

Machine Learning in the abstract

Abstract ML

H	$\mathcal{S}$	X	$\mathcal{L}$
hypotheses space	training input	descriptions of inputs	outputs

Supervised learning

H	$\mathcal{S}$	X	$\mathcal{L}$
h_s	training instances	feature space	classifications
$h_s : X \to \mathcal{L}$	(x, l)	$x \in X$	$l \in \mathcal{L}$
$h_s(x) = l$			

Supervised learning - ABML

H	$\mathcal{S}$	X	$\mathcal{L}$
IF F_1 AND...AND F_n THEN C		$Reasons \subseteq X$ $F_i \in X$	$C \in \mathcal{L}$

Arguments:

- C because Reasons or
- C despite Reasons

AF: ✗

Semantics: ✗

Supervised learning - CLA

H	$\mathcal{S}$	X	$\mathcal{L}$
		$x \in X$	$l \in \mathcal{L}$

Arguments: $\langle h, x, l \rangle$

- if $h = \emptyset$ then $(x, l) \in \mathcal{S}$, and
- if $h \neq \emptyset$ then $h(x) = l$

AF: AA with preferences

Semantics: extensions

Preference relation over H:

- arguments from $\mathcal{S}$are stronger than arguments obtained from H;
- arguments from most preferred hypotheses are stronger than arguments from less preferred hypotheses.

Supervised learning - MAICL

H	$\mathcal{S}$	X	$\mathcal{L}$
		$F_i \in X$	$\mathcal{L} = \{0, 1\}$
			$C \in \mathcal{L}$

Arguments: IF F_1 AND...AND F_n THEN C

AF: AA

Semantics: dialectical trees

Supervised learning - CleAr

$$\frac{H \quad \mathcal{S} \qquad X \qquad \mathcal{L}}{\textit{Premise} \subseteq X \quad \textit{Conclusion} \in L}$$

Arguments: $\textit{Premise} \rightarrow \textit{Conclusion}$

AF: Bipolar AA & QuAD

Semantics: Quantitative

Unsupervised learning

H	$\mathcal{S}$	X	$\mathcal{L}$
h_u	training instances	feature space	obtained from
$h_u : X \to \mathcal{L}$	$\mathcal{S} \subseteq X$		'learnt' clusters

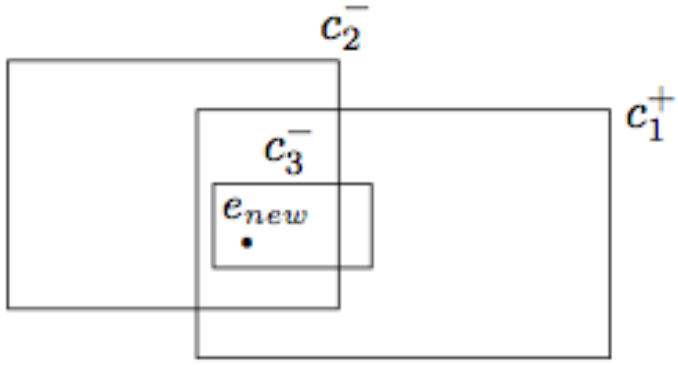
Unsupervised learning - A-ART

Arguments: DeLP

AF: DeLP

Semantics: dialectical trees

Arguments:
+ (e belongs to c_1^+)
- (e belongs to c_3^-)



Reinforcement learning

H	$\mathcal{S}$	X	$\mathcal{L}$
h_r	reward function	states	actions
$h_r : X \to \mathcal{L}$	$\mathcal{S} : X \to R$		

Reinforcement learning - AARL

H	$\mathcal{S}$	X	$\mathcal{L}$
		$Premise \subseteq X$	$Conclusion \in \mathcal{L}$

Arguments: *Conclusion* IF *Premise*

AF: Value-based AA

Semantics: extensions

Comparison of approaches

Comparison of approaches

Method	ML method	AF	Semantics	D/A ML	Multi agent	Advantages	Apps.
ABML	CN2	✗	✗	D		experimental (accuracy, robustness); elicitation	law; medicine; zoology; chess; coding
ABILP	ILP	✗	✗	D			
CLA	concept learning	AA with prefs.	extensions	✗		theoretical (inconsistency tolerance); explanation	
MAICL	concept learning	AA	dialectical trees	A	✓	experimental (recall); partial info	
CleAr	Random Forests; NB; SVM	Bipolar AA/ QuAD	quantitative	A		experimental (accuracy)	Sentiment Analysis; Argument Mining
A-ART	Fuzzy ART	DeLP	dialectical trees	A		explanation; inconsistency resolution	
AARL	SARSA	Value-based AA	extensions	D	✓	experimental (stability; convergence time; optimal performance)	RoboCup; Wumpus

Table 1. Overview of approaches using argumentation to aid ML (D=During, A=After, Apps. = Applications).

Thank you!

Supervised learning - ABML Example

$x = PaysRegularly = no, Rich = yes, HairColor = blond$
$I = CreditApproved$

Arguments:

- *CreditApproved* because $Rich = yes$
- *CreditApproved* despite $PaysRegularly = no$

CN2: IF $HairColor =$ blond THEN *CreditApproved*

ABML: IF $HairColor =$ blond AND $Rich =$ yes THEN *CreditApproved*

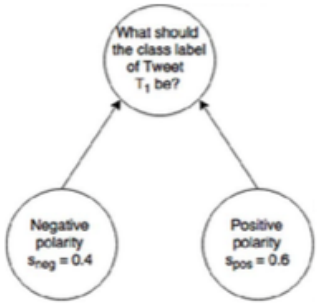
Supervised learning - CleAr Example

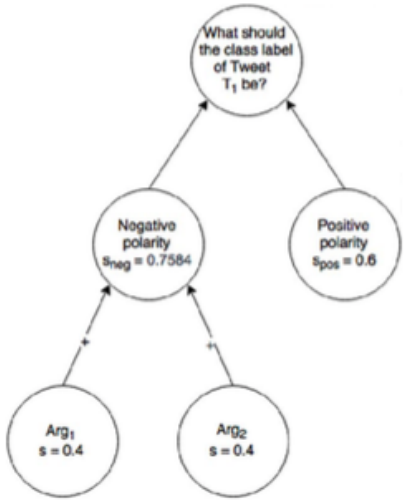
more depressed than you could ever imagine that I wont be going to Vegas. I hate having to be financially responsible

$\mathcal{L} = \{positive, negative\}$

Arguments:

- 'hate' occurs in the tweet $\rightarrow$ negative
- a negation ('wont') occurs in the tweet $\rightarrow$ negative





Reinforcement learning - AARL Example

Arguments:

- agent a_1 should tackle the ball IF a_1 is closest to the ball keeper
- agent a_1 should mark agent a_2 IF a_1 is closest to a_2

If *tackling* is more preferred than *marking* then:

- the attack from the second to the first argument is deleted
- *tackling* gets extra reward at the current iteration of learning