Semantic Relation Extraction from Legislative Text using Generalized Syntactic Dependencies and Support Vector Machines

Guido Boella, Luigi Di Caro, and Livio Robaldo

7th International Web Rule Symposium (RuleML 2013)

Special Track on "Translating between Human Language and Formal Rules: Business, Law, and Government"

Overview

- Introduction
- An example
- Approach
  - Generalization of syntactic contexts
  - Learning step
  - Classification step
- Evaluation
  - Data
  - Results

Introduction

- huge amount of legal data coming from different sources of information
  - how to semantically analyse such data in order to access, reuse, and create knowledge
    - Knowledge acquisition bottleneck

- we propose a novel technique to automatically identify semantic relations in legal text, making use of:
  1. the best approaches for linguistic analyses (POS-tagger, dependency parser)
     - TULE Italian and English parser - (Lesmo L., 2009)
  2. the best approach for standard text classification (Support Vector Machines)
     - Support Vector Machines - (Joachims, T. 1998)
  3. and an easy way to make these two modules communicate

Example

- Legal text
  - *"Under penalty of 2500 to 6400 euros or a three to six months detention, the* ***employer*** *must maintain the personal* ***protective equipment*** *and ensure the* ***hygiene conditions*** *of the* ***employees*** *through maintenance, repairs and replace... necessary and in accordance with any instructio... by the manufacturer"*

pena di una ammenda da 2500 a 6400 euro o dell'arresto da tre a sei mesi, il datore di lavoro deve mantenere in efficienza i dispositivi di protezione individuale e assicurare le condizioni d'igiene, mediante la manutenzione, le riparazioni e le sostituzioni necessarie e secondo le eventuali indicazioni fornite dal fabbricante.

- Ser...
  - *A...*
  - *P...*
  - *I...*
  - *I...*

Approach

- Given a set of semantic annotations *sem*(x_i) between a noun x_i and a semantic tag *sem*
  - extract all syntactic contexts of each x_i
  - extract all syntactic contexts of the other nouns y_j
  - label all the x_i as positive examples (for the semantic tag *sem*)
  - label all the y_i as negative examples (for the semantic tag *sem*)
  - generalize all syntactic contexts
  - learn a *sem*-model using a Support Vector Machine

- When parsing a new text, all the syntactic contexts of every noun are passed to the *sem*-model
  - The classifier will decide if they can be annotated with the semantic tag *sem*, one by one

Generalization of syntactic contexts

- Given a noun x_i
  - extract all the syntactic dependencies *dep*(x_i, k) *or dep*(k, x_i) of x_i .
  - substitute all involved terms k with the generic string "*noun*" in case they are nouns. Otherwise, leave them as they are. This creates a generalization over the nouns that are syntactically linked to the x_i
  - substitute x_i with the string "*target*" to generalize its position (as a left or right argument)
  - form a single token that represents the syntactic information unit for each syntactic dependency
    - Example:

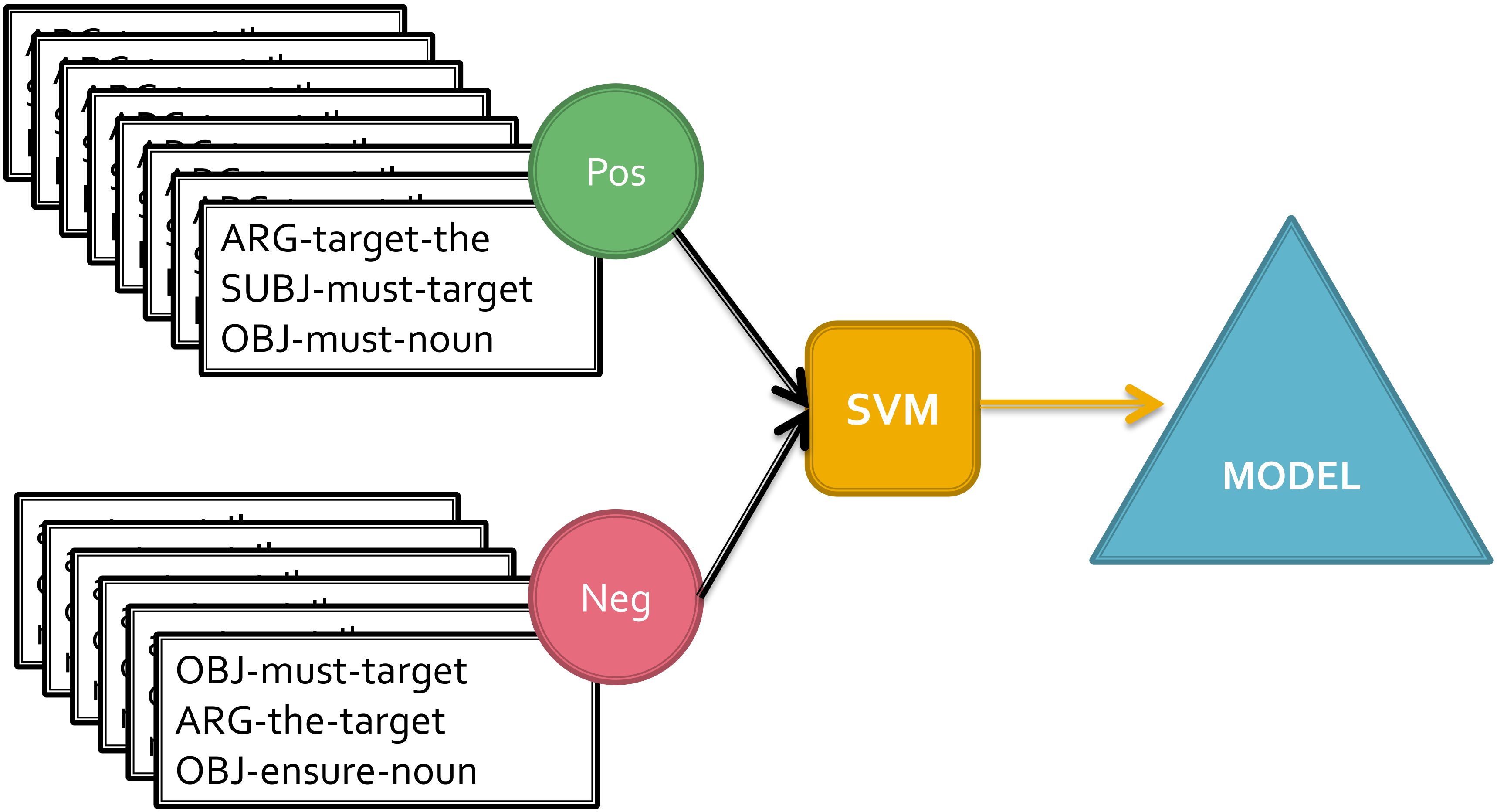
rmod(x_i , k) → rmod-target-noun

Example (for noun "employer")

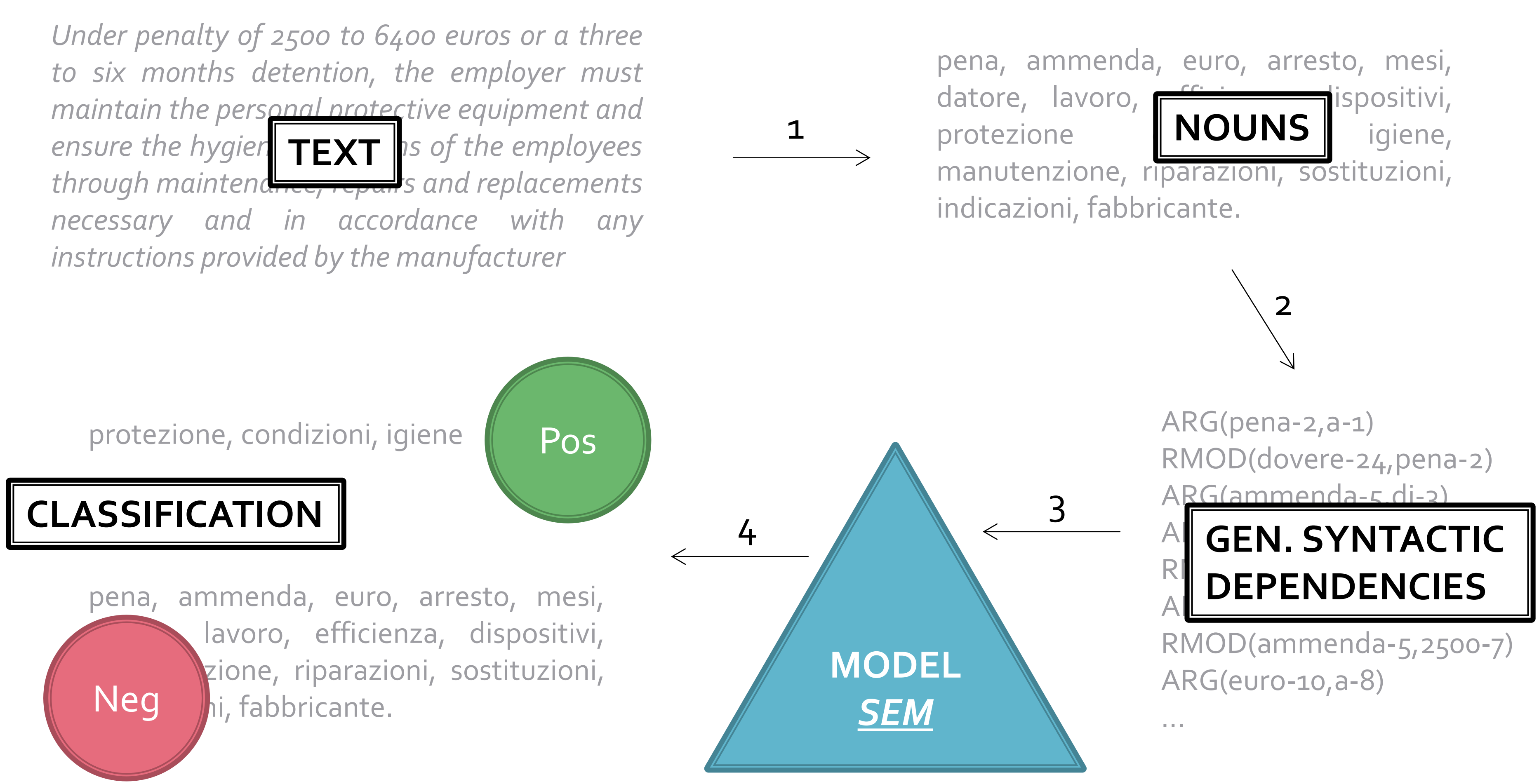
- Legal text
  - *"Under penalty of 2500 to 6400 euros or a three to six months detention, the* ***employer*** *must maintain the personal protective equipment and ensure the hygiene conditions of the employees through maintenance, repairs and replacements necessary and in accordance with any instructions provided by the manufacturer"*.

ARG(employer, the) SUBJ(must, employer) OBJ(must, equipment)	→	ARG-target-the SUBJ-must-target OBJ-must-noun
Syntactic context for noun employer		**<u>Generalized</u> syntactic context for noun "employer"**

Learning step



Classification step



Evaluation

- Dataset
  - It contain 156 legal texts annotated with semantic information, containing a total of 2253 nouns.

- Tested semantic annotations:
  - **Active role.** The active role indicates an active agent involved within the situation described in the text. Examples of common entities related to active roles are employers, directors of banks, doctors, security managers.
  - **Passive role.** The passive role indicates an agent that is the beneficiary of the described norm. Examples of agents associated with passive roles are workers and work supervisors.
  - **Involved Object.** An involved object represents an entity that is central for the situation being described. Examples are types of risk for a worker, the location of a specific work, and so on.

Evaluation (2)

- Results

Active Role	*Precision*	*Recall*	*F-Measure*
yes	97.2%	92.6%	94.8%
no	99.3%	99.8%	99.5%

Passive Role	*Precision*	*Recall*	*F-Measure*
yes	100.0%	26.8%	42.3%
no	98.7%	100.0%	99.3%

Involved Object	*Precision*	*Recall*	*F-Measure*
yes	59.3%	31.9%	41.4%
no	91.3%	97.0%	94.1%

Thank you

Questions?