Distant supervised learning for the TAC-KBP Slot Filling and Temporal Slot Filling Tasks

Guillermo Garrido, Bernardo Cabaleiro, Anselmo Peñas, Álvaro Rodrigo, Damiano Spina
NLP & IR Group at UNED
{
(ggarrido, bcbaleiro, anselmo, alvarory, damiano)@lsi.uned.es

1. A distant supervised system, specialized for the Regular Slot Filling and Full Temporal Slot Filling subtasks
   - Automatically gather training data for supervised slot classifiers from the initial Knowledge Base and source document collection.
   - Use a rich document representation, augmenting syntactic dependency trees.
   - Find and aggregate time constraints for the same slot value across different documents.

2. System Description: distant supervised learning.


4. Slot temporal restrictions.
   - Use Tarso to get temporal relations: included, simultaneous, after, before, begun_by, ended.
   - Temporal relations acquisition: identify a syntactic pattern: `<EVENT> - <PREPOSITION> - <TIME EXPRESSION>`
   - Transform into one of: within, throughout, beginning, ending, after, before.

5. Learning extractors.
   - Gathering of distant training examples: from a seed triple `<entity, slot type, value>`
     - we retrieve candidate documents that contain both entity and value.
   - Named Entity type matching.
   - Each example is represented by binary features.
   - Classification process: supervised classification (linear SVM).
   - Answer aggregation

6. Regular Slot Filling Subtask (SF).
   - By the time of submission, the system was not fully developed (we do not report results in this poster).
   - Our training did not cover all seeds: use of supervised seeds.
   - SVM multi-class classifier with the positive and negative examples.
   - Results below average of the systems.

7. Temporal Slot Filling Subtask (TSF).
   - We used a battery of binary classifiers: SVMlight.
   - Once extracted the `<entity, slot type, value >`, temporal constraints are generated depending of semantics of the event, slot type and the temporal restriction found.
   - Generated temporal constraints are aggregated.
   - Results slightly above the median of the systems.

8. Preliminary Results: 2011 Temporal SF full task scores

   - The performance of our simple distant learning system varies by slot type.
   - Our systems (TSF) have the highest precision among participants, but low recall.
   - Graph representation has helped: we expect a performance improvement from a better document representation.
   - Simple aggregation of dates found in documents was a strong baseline we could not beat.