Slot Filler Validation at TAC 2014
Task Guidelines
Version 1.0, April 17, 2014

Introduction

The goal of Knowledge Base Population (KBP) at TAC 2014 is to promote research in and to evaluate the ability of automated systems to discover information about named entities and incorporate this information in a knowledge source. For the evaluation an initial (or reference) knowledge base (KB) is provided along with a source document collection from which systems are to discover information to populate the reference KB. Attributes (a.k.a., "slots") derived from Wikipedia Infoboxes have been used to create the reference knowledge base. The overall task of populating a knowledge base is decomposed into two major tasks: Entity Linking (EL), where names must be aligned to entities; and Slot Filling (SF), which involves mining information about entities from text. Another track that has been proposed in KBP is the Slot Filler Validation (SFV) Track. SFV is a diagnostic task that eliminates the need for a team to have a full Slot Filling system and aims at validating the output of the systems participating in the KBP Slot Filling track. For the other tracks in KBP 2014, please visit the KBP 2014 web page: http://www.nist.gov/tac/2014/KBP/.

A variant of the Slot Filler Validation task was originally proposed in the Recognizing Textual Entailment (RTE) track of TAC 2010 and TAC 2011, which aimed to show the potential utility of RTE systems for Knowledge Base Population. In the RTE KBP Validation task, the "text" consisted of the provenance (document) returned with each candidate slot filler, and the "hypothesis" was a natural language expression of the relation between the target entity and the candidate filler; the text entailed the hypothesis if and only if the candidate slot filler was Correct according to the associated provenance. The input to the RTE system was the set of T-H pairs generated from the pooled results of all SF systems, and the evaluation metric was P/R/F1 on the pooled SF output.

In TAC 2013, the Slot Filler Validation track was proposed with a slightly different use case and evaluation metric than those in RTE-6 and RTE-7. The 2013 SFV track focused on the refinement of output from slot filling systems by either applying more intensive linguistic processing to validate individual candidate slot fillers (as in RTE-6 and RTE-7) *or combining information from multiple slot filling systems*. Evaluation was based on the change in P/R/F1 of individual SF systems, rather than on P/R/F1 of the single pool of SF output.

Slot Filler Validation is again proposed at KBP 2014 without significant changes to SFV at KBP 2013. This document provides a definition of the KBP 2014 Slot Filler Validation Task and a description of the data set, together with instructions on how to take part in the exercise. **It is assumed that the reader has already read the task definition for the KBP 2014 English slot filling track (http://surdeanu.info/kbp2014/def.php).**

Task Description

The slot filler validation (SFV) track is motivated by a use case in which the SFV system is used as a component of a full SF system. It focuses on the refinement of output from English slot filling (SF) systems by either combining information from multiple slot filling systems, or applying more intensive linguistic processing to validate individual candidate slot fillers.

The KBP *Slot Filling* task is focused on searching a collection of newswire, Web, and discussion forum documents and extracting values for a pre-defined set of attributes ("slots") for target entities. Given an entity in a knowledge base and an attribute for that entity, systems must find in a large corpus the correct value(s) for that attribute and return the slot filler together with its provenance, where provenance is a set of text spans from documents in the corpus that justify the correctness of the slot filler.

The input to the *Slot Filler Validation* system is a set of submission files from several slot-filling runs (with the run ID anonymized appropriately). The output of the SFV system is a single tab-delimited file with binary classifications (Correct/Incorrect, i.e., 1/-1) of each candidate slot filler in each SF run.

The evaluation measures the effect of using the SFV output to filter the contributing SF runs. Each contributing SF run will be filtered and re-scored in the same way as for the full SF task, and the results compared against the scores for the unfiltered SF run. SFV tries to increase Precision of the contributing SF runs without significantly reducing Recall, and the objective function is to maximize the mean F-score over each of the filtered contributing runs.

Data

The Slot Filler Validation data set is based on the runs submitted in response to the TAC 2014 KBP English Slot Filling task.

1   Input Format

The input to the slot filler validation task consists of 4 types of files:
   1)   SF input (SF query file, reference KB, source documents)
   2)   SF system output files, generated in response to SF input
   3)   SF System profile file (optional)
   4)   Assessment of all SF system output for a small number of SF query+slot combinations (optional)

The slot definitions, assessment guidelines, and description of the Slot Filling input and output are available on the TAC 2014 English Slot Filling website at
http://surdeanu.info/kbp2014/def.php.

For the slot-filler validation task, the KBP 2014 SF ***teams*** will be anonymized and renamed SFV_01, SFV_02, SFV_03,..., SFV_XX. The output files of the SF runs will have the run ID in Column 3 replaced by a SFV query ID consisting of the anonymized team_name + run_number

+ filler_candidate_number. Thus, the SFV query ID shows which slot filler candidates came from the same SF run and which SF runs were produced by the same team, for the benefit of SFV teams who may want an idea of system diversity when applying cross-system voting approaches.

2   Output Format

The output of a SFV run should be a single tab-delimited file with exactly one judgment for each SFV query ID. The possible judgments are:

  1: Candidate slot filler is Correct or Redundant with reference KB
 -1: Candidate slot filler is Wrong or Inexact

The definitions of Correct, Redundant, Wrong, and Inexact for slot fillers are given in the TAC KBP 2014 English Slot Filling task description at http://surdeanu.info/kbp2014/def.php.

The coordinators intend to compute another score like RTE-6 KBP Validation score for TAC KBP-2014. To facilitate that addition, SFV teams are required to produce the same decision (classification) for candidate answers that have the same slot filler string and provenance, regardless of which SF run or SF team they come from. The new scoring type will be added in future versions of this guideline.

3   Training Data

Evaluation data from the KBP 2013 SFV track will be provided to SFV 2014 participants to facilitate system development. The training package will be available on the TAC KBP 2014 website at http://www.nist.gov/tac/2014/KBP/data.html.

Scoring and Metrics

System results will be compared to the human-annotated gold standard and the metrics used to evaluate system performances will be F-measure and their difference with the original unfiltered SF runs F-measure. Other measures are under investigation and will be announced in the next versions of this guideline.

Submissions

Participants will have one week after the evaluation queries are released to return their results. Up to five alternative system runs may be submitted by each team. Submitted runs should be ranked according to their expected score (based on development data, for example). Systems should not be modified once queries are downloaded. Details about submission procedures will be communicated to the track mailing list.   The   tools   to   validate   formats   will   be   made   available   at: http://www.nist.gov/tac/2014/KBP/SFValidation/tools.html

Schedule

Please visit the KBP 2014 slot filler validation webpage for updates on schedule: http://www.nist.gov/tac/2014/KBP/SFValidation/index.html

March: Track guidelines posted
April: Distribution of document collections
June 15: Deadline for registration for track participation
September 2-9: Slot Filler Validation evaluation window
By late September: Release of individual evaluated results to participants
October 7: Deadline for short system descriptions
October 7: Deadline for workshop presentation proposals
Mid October: Notification of acceptance of presentation proposals
November 1: Deadline for system reports (workshop notebook version)
November 17-18: TAC 2014 workshop in Gaithersburg, Maryland, USA
February 15, 2015: Deadline for system reports (final proceedings version)

Mailing List and Website

The Slot Filler Validation track website is http://www.nist.gov/tac/2014/KBP/SFValidation/. The KBP mailing list is tac-kbp@nist.gov. Information about subscribing to the list is available at: http://www.nist.gov/tac/2014/KBP/registration.html.