The Annotation Guideline for Syntactic Relations among Eligibility Criteria

April 2016
Tian Kang

Overall Goal

To develop a machine learning-based parser for clinical trial eligibility criteria texts; use the parser to extract concepts and their clinical relations from the free-text eligibility criteria and re-represent it in a structured way.

Subtasks

1. Named Entity Recognition (NER)
2. **Syntactic relations identification**
3. Negation/Hedge detection
4. Semantic relations identification

This guideline is designed for the second task: **Syntactic relations identification**, i.e., identification of the relations between concepts from entity classes (Condition, observation, drug/substance, device/procedure) and attributes (qualifier/modifier, temporal constrains, measurement, anatomic location) (details about concepts see concept annotation guideline).

Introduction to this guideline

In this task, we define the relations between the entities and attributes based on the concept annotation results of NER task. By definition "Relations are triples, containing two concepts, and either a relation strength (unnamed relations) or a relation name (named relations). Out of these triples we build a network whose nodes are entities and concepts and whose arcs, connecting the nodes, are the relationships." In this step, we only focus on **syntactic relations between the entities and attributes**, which can be clearly understood by only analyzing the syntax of the sentence without requiring any domain knowledge.

Data

Annotators will apply this annotation guideline to clinical trial eligibility criteria texts retrieved from ClinicalTrials.gov to make the training data for machine learning-based NER task.

Contents in this guideline:

2. General annotation rules – p4

Define a set of basic rules that need to be followed the entire annotation process and include annotation guidelines for some specific cases to avoid ambiguity

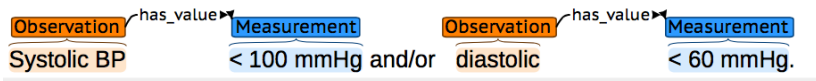
*****Example criteria texts or concepts in this guideline are all *italic (all examples are from clinicaltrials.gov on neuropsychological diseases).*

Overview of the syntactic relation categories

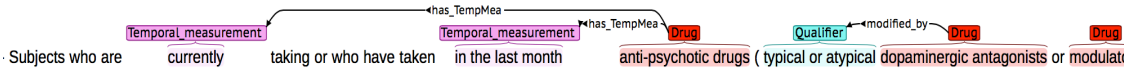
Relation	Concept Categories direction		Definition	Examples
	Left	Right		
Has_value	Condition Observation Procedure/Device Drug/Substance	Measurement	Link entities to the their numerical measurement	*-Omega- 3 fatty acids (> 1000 mg/day)* Relations*:* *Omega- 3 fatty acids has_value > 1000 mg/day*
Has_TemMea	Condition Observation Procedure/Device Drug/Substance	Temporal constraints	Link entities to the their temporal constraints	*-Stability of Permitted Medications for 4 weeks prior to the enrollment.* Relations*:* *Permitted Medication has_TemCon 4 weeks*
Modified by	Condition Observation Procedure/Device Drug/Substance	Qualifier/Modifier	Link entities to the their pre- post-coordination.	*-Typical or atypical neuroleptics* Relations: *Neuroleptics modified by Typical or atypical* *-Stability of Permitted Medications for 4 weeks prior to the enrollment.* Relations*:* *Permitted Medication has_TemCon 4 weeks* *Permitted Medication modified by Stability*
Located_in	Condition Observation Procedure/Device Drug/Substance	Anatomic location	Link entities to the their anatomic locations.	*-Participants with multiple lacunes or lacunes in a critical memory structure are excluded* Relations: *Lacunes Located_in critical memory structure*

General annotation rules:

1. **Relation direction:** relation direction is always from an entity to its attributes

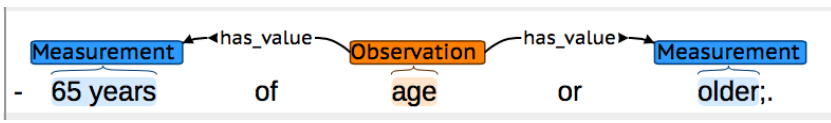


2. **When an entity with attributes has multiple specified examples, only link the attribu with general entity.**



3. **We do not annotate the relations within entity classes (requiring domain knowledge, e. condition 1 *caused by* condition 2. This task is "semantic relation identification")**

4. **When an attribute is separated by other concepts, link the entity to its all separat segments.**



5. **Rules of drug dosage annotation (when concerning the status of the dosage)**

Dosage modified by dosage status

Dosage has temporal constraints

(relation between dosage and drug is not included in this tasks, but included in a semantic task)