

Master in Artificial Intelligence

Goals

The Task:
SemEval 2013

Dataset

Course Plan

Advanced Human Language Technologies



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona



Outline

Goals

The Task:
SemEval 2013

Dataset

Course Plan

1 Goals

2 The Task: SemEval 2013

3 Dataset

4 Course Plan

Lab Project Goals

Goals

The Task:
SemEval 2013

Dataset

Course Plan

- Experiment with different AI approaches for a domain-specific NLP task.
- The goal **is not** to build the best possible system, but to understand and compare different NLP paradigms.

Outline

Goals

The Task:
SemEval 2013

Dataset

Course Plan

1 Goals

2 The Task: SemEval 2013

3 Dataset

4 Course Plan

SemEval-2013 Task 9: DDI Extraction

Goals

The Task:
SemEval 2013

Dataset

Course Plan

- **SemEval** - International Conference on Semantic Evaluation. Several tasks or challenges are posed every edition
- **SemEval 2013 Task 9: DDIExtraction** - Detect drug names and interactions among them described in text.
- **Documents** - Documents extracted from DrugBank (Drug description leaflets database) and MedLine (abstracts of medical papers)
- **Participants** - 11 research teams from around the world (1 Cuba, 1 Italy, 1 Finland, 2 Germany, 1 Portugal, 2 Spain, 3 USA)
- **Subtasks**
 - Drug name recognition and classification (NERC)
 - Drug-Drug interaction recognition and classification (DDI)

Outline

Goals

The Task:
SemEval 2013

Dataset

Course Plan

1 Goals

2 The Task: SemEval 2013

3 Dataset

4 Course Plan

DDI Extraction Dataset

DrugBank document example

Goals

The Task:
SemEval 2013

Dataset

Course Plan

```
<document id="DDI-DrugBank.d483">
  <sentence id="DDI-DrugBank.d483.s0" text="Increases in serum creatinine have occurred when CEFOTAN was given alone.">
    <entity id="DDI-DrugBank.d483.s0.e0" charOffset="49-55" type="brand" text="CEFOTAN"/>
  </sentence>
  <sentence id="DDI-DrugBank.d483.s1"
    text="If CEFOTAN and an aminoglycoside are used concomitantly, renal function should be carefully monitored,
    because nephrotoxicity may be potentiated.">
    <entity id="DDI-DrugBank.d483.s1.e0" charOffset="3-9" type="brand" text="CEFOTAN"/>
    <entity id="DDI-DrugBank.d483.s1.e1" charOffset="18-31" type="group" text="aminoglycoside"/>
    <pair id="DDI-DrugBank.d483.s1.p0" e1="DDI-DrugBank.d483.s1.e0" e2="DDI-DrugBank.d483.s1.e1" ddi="true" type="effect"/>
  </sentence>
  (...)
  <sentence id="DDI-DrugBank.d483.s4"
    text="As with other cephalosporins, high concentrations of cefotetan may interfere with measurement of
    serum and urine creatinine levels by Jaffe reaction and produce false increases in the levels
    of creatinine reported.">
    <entity id="DDI-DrugBank.d483.s4.e0" charOffset="14-27" type="group" text="cephalosporins"/>
    <entity id="DDI-DrugBank.d483.s4.e1" charOffset="53-61" type="drug" text="cefotetan"/>
    <pair id="DDI-DrugBank.d483.s4.p0" e1="DDI-DrugBank.d483.s4.e0" e2="DDI-DrugBank.d483.s4.e1" ddi="false"/>
  </sentence>
</document>
```

DDI Extraction Dataset

MedLine document example

Goals

The Task:
SemEval 2013

Dataset

Course Plan

```
<document id="DDI-MedLine.d16">
  <sentence id="DDI-MedLine.d16.s0"
    text="Interaction of ketamine and halothane in rats.">
    <entity id="DDI-MedLine.d16.s0.e0" charOffset="15-22" type="drug" text="ketamine"/>
    <entity id="DDI-MedLine.d16.s0.e1" charOffset="28-36" type="drug" text="halothane"/>
    <pair id="DDI-MedLine.d16.s0.p0" e1="DDI-MedLine.d16.s0.e0" e2="DDI-MedLine.d16.s0.e1" ddi="true" type="int"/>
  </sentence>
  <sentence id="DDI-MedLine.d16.s1"
    text="The interaction of intramuscularly injected ketamine and its N-demethylated metabolite (metabolite I)
    with halothane was evaluated in rats. ">
    <entity id="DDI-MedLine.d16.s1.e0" charOffset="44-51" type="drug" text="ketamine"/>
    <entity id="DDI-MedLine.d16.s1.e1" charOffset="107-115" type="drug" text="halothane"/>
    <pair id="DDI-MedLine.d16.s1.p0" e1="DDI-MedLine.d16.s1.e0" e2="DDI-MedLine.d16.s1.e1" ddi="false"/>
  </sentence>
  (...)
  <sentence id="DDI-MedLine.d16.s5"
    text="The half-life of ketamine in plasma and brain was longer in the presence of halothane than when
    ketamine was given alone. ">
    <entity id="DDI-MedLine.d16.s5.e0" charOffset="17-24" type="drug" text="ketamine"/>
    <entity id="DDI-MedLine.d16.s5.e1" charOffset="76-84" type="drug" text="halothane"/>
    <entity id="DDI-MedLine.d16.s5.e2" charOffset="96-103" type="drug" text="ketamine"/>
    <pair id="DDI-MedLine.d16.s5.p0" e1="DDI-MedLine.d16.s5.e0" e2="DDI-MedLine.d16.s5.e1" ddi="true" type="mechanism"/>
    <pair id="DDI-MedLine.d16.s5.p1" e1="DDI-MedLine.d16.s5.e0" e2="DDI-MedLine.d16.s5.e2" ddi="false"/>
    <pair id="DDI-MedLine.d16.s5.p2" e1="DDI-MedLine.d16.s5.e1" e2="DDI-MedLine.d16.s5.e2" ddi="false"/>
  </sentence>
  (...)
</document>
```


Outline

Goals

The Task:
SemEval 2013

Dataset

Course Plan

1 Goals

2 The Task: SemEval 2013

3 Dataset

4 Course Plan

Course plan for lab sessions

Goals

The Task:
SemEval 2013

Dataset

Course Plan

- Sessions 1-2: Experiment and improve ML-based drug NERC
- Sessions 3-4: Experiment and improve NN-based drug NERC
- Sessions 5-7: Experiment and improve LLM-based drug NERC
- **Deliverable:** Report on NERC
- Sessions 8-9: Experiment and improve ML-based drug DDI
- Session 10: Experiment and improve NN-based drug DDI
- Sessions 11-13: Experiment and improve LLM-based drug DDI
- **Deliverable:** Report on DDI

Course plan for lab sessions

Goals

The Task:
SemEval 2013

Dataset

Course Plan

- Lab project must be developed in teams of 2 people
- **Warning:** Lab sessions may not be enough. You'll probably need to work on the assignments at home.