Introduction to CodaLab Competitions

Tristan Miller

Presented at:
School of Data Analysis and Artificial Intelligence
National Research University – Higher School of Economics
25 May 2017
Overview

What are shared tasks?

What is CodaLab Competitions?

Organizing CodaLab Competitions

Student competitions

Caveats
What are shared tasks?
Introduction

In a shared task / challenge / competition / evaluation campaign / evaluation exercise, the organizers

- define a data processing task (classification, segmentation, ranking, etc.)
Introduction

In a shared task / challenge / competition / evaluation campaign / evaluation exercise, the organizers

▶ define a data processing task (classification, segmentation, ranking, etc.)
▶ define evaluation metrics to measure performance
Introduction

In a shared task / challenge / competition / evaluation campaign / evaluation exercise, the organizers

- define a data processing task (classification, segmentation, ranking, etc.)
- define evaluation metrics to measure performance
- produce test data and the gold standard answers
Introduction

In a shared task / challenge / competition / evaluation campaign / evaluation exercise, the organizers
- define a data processing task (classification, segmentation, ranking, etc.)
- define evaluation metrics to measure performance
- produce test data and the gold standard answers
- solicit participants to write algorithms to process the test data
In a shared task / challenge / competition / evaluation campaign / evaluation exercise, the organizers

- define a data processing task (classification, segmentation, ranking, etc.)
- define evaluation metrics to measure performance
- produce test data and the gold standard answers
- solicit participants to write algorithms to process the test data
- applying the evaluation metrics, score the algorithms’ output against the gold standard
Introduction

In a shared task / challenge / competition / evaluation campaign / evaluation exercise, the organizers

- define a data processing task (classification, segmentation, ranking, etc.)
- define evaluation metrics to measure performance
- produce test data and the gold standard answers
  - produce trial data for demonstration purposes
  - produce or provide ancillary resources (knowledge bases, etc.)
  - produce training data for use by supervised algorithms
- solicit participants to write algorithms to process the test data

- applying the evaluation metrics, score the algorithms’ output against the gold standard
Introduction

In a shared task / challenge / competition / evaluation campaign / evaluation exercise, the organizers

▶ define a data processing task (classification, segmentation, ranking, etc.)
▶ define evaluation metrics to measure performance
▶ produce test data and the gold standard answers
  ▶ produce trial data for demonstration purposes
  ▶ produce or provide ancillary resources (knowledge bases, etc.)
  ▶ produce training data for use by supervised algorithms
▶ solicit participants to write algorithms to process the test data
  ▶ implement various baseline algorithms
▶ applying the evaluation metrics, score the algorithms’ output against the gold standard
Introduction

In a **shared task / challenge / competition / evaluation campaign / evaluation exercise**, the organizers

- define a data processing **task** (classification, segmentation, ranking, etc.)
- define **evaluation metrics** to measure performance
- produce **test data** and the **gold standard** answers
  - produce **trial data** for demonstration purposes
  - produce or provide **ancillary resources** (knowledge bases, etc.)
  - produce **training data** for use by supervised algorithms
- solicit participants to write **algorithms** to process the test data
  - implement various **baseline algorithms**
- applying the evaluation metrics, **score** the algorithms’ output against the gold standard
  - compare and analyze the participants’ and baseline algorithms
Shared tasks: pros and cons

Pros:
- stimulate methodological research on unsolved problems
- provide standardized data sets, resources, and evaluation metrics
- facilitate reproducibility of research results
- centralize publication and discussion of research results

Cons:
- everything must be planned in advance
- large organizational overhead (data distribution, publicity, communication with participants, etc.)
- encourages “teaching to the test”
Shared tasks: pros and cons

Pros:
▶ stimulate methodological research on unsolved problems
▶ provide standardized data sets, resources, and evaluation metrics
▶ facilitate reproducibility of research results
▶ centralize publication and discussion of research results

Cons:
▶ everything must be planned in advance
▶ large organizational overhead (data distribution, publicity, communication with participants, etc.)
▶ encourages “teaching to the test”
What is CodaLab Competitions?
CodaLab Competitions

- Web-based platform for running online data-based competitions
- Developed by Microsoft, Stanford University, and others
- Free hosted implementation: https://competitions.codalab.org/
- Free software (Apache License 2.0): https://github.com/codalab/codalab-competitions/
Microsoft COCO Image Captioning Challenge

The man at bat readies to swing at the pitch while the umpire looks on.

A large bus sitting next to a very tall building.
ChaLearn Looking at People Challenges
SemEval-2017 Multilingual and Cross-lingual Semantic Word Similarity
CodaLab features

▶ Hosts public task website
▶ Hosts private gold-standard data and scoring software
▶ Manages participant registration
▶ Enforces submission deadlines
▶ Runs scoring software
▶ Tabulates, stores, and publishes results
▶ Handles communication between organizers/participants (e-mail, forums)
▶ Provides some publicity
<table>
<thead>
<tr>
<th>Competition</th>
<th>Date</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Story Cloze Test</td>
<td>Aug 30, 2016</td>
<td>30</td>
</tr>
<tr>
<td>Organized by ROCNLP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A challenge for evaluating a system's natural language and story understanding.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Large Scale Movie Description Challenge (LSMDC) 2016 : Movie Multiple-Choice Test</td>
<td>Aug 25, 2016</td>
<td>18</td>
</tr>
<tr>
<td>Organized by atousa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The goal of the challenge is to evaluate different visual-language models performance to annotate videos based on natural sentences for ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Large Scale Movie Description Challenge (LSMDC) 2016 : Movie Retrieval</td>
<td>Aug 25, 2016</td>
<td>17</td>
</tr>
<tr>
<td>Organized by atousa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The goal of the challenge is to evaluate different visual-language models performance to retrieve videos based on natural sentence queries ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organized by julsal</td>
<td>Feb 04, 2017</td>
<td></td>
</tr>
<tr>
<td>End-User Development using Natural Language</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Overview

This is the CodaLab Competition for Subtask 3 of SemEval-2017 Task 7: Detection and Interpretation of English Puns.

Background

A pun is a form of wordplay in which one signifier (e.g., a word or phrase) suggests two or more meanings by exploiting polysemy, or phonological similarity to another signifier, for an intended humorous or rhetorical effect. For example, the first of the following two punning jokes exploits contrasting meanings of the word "interest", while the second exploits the sound similarity between the surface form "propane" and the latent target "profane":

I used to be a banker but I lost interest.

When the church bought gas for their annual barbecue, proceeds went from the sacred to the propane.
SemEval-2017 Task 7, Subtask 3
Organized by Logological - Current server time: May 7, 2017, 2:11 p.m. UTC

Evaluation criteria

The evaluation for this subtask will be carried out in two simultaneous phases, one for the homographic data set and one for the heterographic data set. Systems may participate in either or both phases.

Systems participating in a given phase may provide a single guess for any or all of the contexts in the data set.

The results for each phase must be submitted in a delimited text file named `answer.txt`. Each line of the text file consists of three fields separated by horizontal whitespace (a single tab or space character). The first field is the ID of a pun word from the data set. The second field is a semicolon-delimited list of WordNet 3.1 sense keys that match one meaning of the pun. The third field is a semicolon-delimited list of WordNet 3.1 sense keys that match the other meaning of the pun. Sample data and results files are available in the trial data.

To submit the results, place `answer.txt` in a ZIP file (in the top-level directory), and then upload it to CodaLab according to the instructions at Participating in a Competition.
SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:11 p.m. UTC

**Current**
- **Trial**
  - Dec. 5, 2016, midnight UTC

**Next**
- **Test (Homographic)**
  - Jan. 23, 2017, midnight UTC

**Learn the Details**
- Phases
- Participate
- Results
- Forums

**Trial**
- **Start:** Dec. 5, 2016, midnight

**Test (Homographic)**
- **Start:** Jan. 23, 2017, midnight

**Test (Heterographic)**
- **Start:** Jan. 23, 2017, midnight

**Competition Ends**
- Jan. 30, 2017, 11:59 p.m.
SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

<table>
<thead>
<tr>
<th>Current</th>
<th>Next</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>Test (Homographic)</td>
</tr>
</tbody>
</table>

Get Data
Submit / View Results

Data sets

Data sets for this subtask are described in detail on the SemEval-2017 Task 7 website.

Download

- Trial and test data

Note that, due to the difficulty in amassing a large number of pun examples per word or per sense, there is **no training data** for this task.
Phase description
None

Max submissions per day: 999
Max submissions total: 999

Click the Submit button to upload a new submission.

Optionaly add more information about this submission

Submit

Here are your submissions to date (✔ indicates submission on leaderboard):

<table>
<thead>
<tr>
<th>#</th>
<th>FILENAME</th>
<th>SUBMISSION DATE</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>allwrong.zip</td>
<td>12/19/2016 13:06:30</td>
<td>Finished</td>
</tr>
<tr>
<td>2</td>
<td>also_correct.zip</td>
<td>12/19/2016 13:06:32</td>
<td>Finished</td>
</tr>
<tr>
<td>3</td>
<td>badfields.zip</td>
<td>12/19/2016 13:06:35</td>
<td>Failed</td>
</tr>
<tr>
<td>4</td>
<td>badid.zip</td>
<td>12/19/2016 13:06:37</td>
<td>Failed</td>
</tr>
<tr>
<td>5</td>
<td>correct.zip</td>
<td>12/19/2016 13:06:40</td>
<td>Finished</td>
</tr>
<tr>
<td>6</td>
<td>clup.zip</td>
<td>12/19/2016 13:06:43</td>
<td>Failed</td>
</tr>
<tr>
<td>7</td>
<td>halfcoverage.zip</td>
<td>12/19/2016 13:06:45</td>
<td>Finished</td>
</tr>
</tbody>
</table>
SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

**Current**  
- **Trial**  
  - Dec. 5, 2016, midnight UTC

**Next**  
- **Test (Homographic)**  
  - Jan. 23, 2017, midnight UTC

---

### Phase description

None

### Max submissions per day: 999

### Max submissions total: 2

---

### Results

<table>
<thead>
<tr>
<th>#</th>
<th>User</th>
<th>Team Name</th>
<th>F1 ▲</th>
<th>recall ▲</th>
<th>precision ▲</th>
<th>coverage ▲</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>doogyb</td>
<td></td>
<td>0.0774 (4)</td>
<td>0.0770 (4)</td>
<td>0.0778 (4)</td>
<td>0.9900 (1)</td>
</tr>
<tr>
<td>2</td>
<td>kevang</td>
<td>BuZzSw</td>
<td>0.1544 (2)</td>
<td>0.1525 (1)</td>
<td>0.1563 (2)</td>
<td>0.9761 (2)</td>
</tr>
<tr>
<td>3</td>
<td>lhurtado</td>
<td>ELiRF-UPV</td>
<td>0.0996 (3)</td>
<td>0.0976 (3)</td>
<td>0.1014 (3)</td>
<td>0.9646 (3)</td>
</tr>
<tr>
<td>4</td>
<td>evrog</td>
<td>UTMN</td>
<td>0.0452 (5)</td>
<td>0.0424 (5)</td>
<td>0.0484 (5)</td>
<td>0.8760 (4)</td>
</tr>
<tr>
<td>5</td>
<td>tpederse</td>
<td></td>
<td>0.1557 (1)</td>
<td>0.1448 (2)</td>
<td>0.1683 (1)</td>
<td>0.8606 (5)</td>
</tr>
</tbody>
</table>
Organizing CodaLab Competitions
<table>
<thead>
<tr>
<th>Competition Name</th>
<th>Start Date</th>
<th>End Date</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>SemEval-2017 Task 7, Subtask 3</td>
<td>Dec 05, 2016</td>
<td>Jan 30, 2017</td>
<td>15</td>
</tr>
<tr>
<td>SemEval-2017 Task 7, Subtask 2</td>
<td>Dec 05, 2016</td>
<td>Jan 22, 2017</td>
<td>16</td>
</tr>
<tr>
<td>SemEval-2017 Task 7, Subtask 1</td>
<td>Dec 05, 2016</td>
<td>Jan 15, 2017</td>
<td>16</td>
</tr>
</tbody>
</table>
Create Competition

To create a competition, upload a bundle with the definition of the new competition.

Upload definition...
To create a competition, upload a bundle with the definition of the new competition.

Upload definition...
To create a competition, upload a file with the definition of the competition:

```
---
title: SemEval-2017 Task 7, Subtask 3
description: Interpretation of English Puns
image: semeval2017-task7-logo.png
has_registration: True
allow_teams: True
end_date: 2017-01-30
html:
  overview: overview.html
evaluation: evaluation.html
terms: terms_and_conditions.html
data: data.html
phases:
  1:
    phasename: 1
    label: "Trial"
    start_date: 2016-12-05
    max_submissions: 999
    scoring_program: scorer.zip
    reference_data: data_trial.zip
    leaderboard_management_mode: hide_results
    color: white
  2:
    phasename: 2
    label: "Test (Homographic)"
    start_date: 2017-01-23
    max_submissions: 2
    scoring_program: scorer.zip
    reference_data: data_hom.zip
    color: blue
  3:
    phasename: 3
    label: "Test (Heterographic)"
    start_date: 2017-01-23
    max_submissions: 2
    scoring_program: scorer.zip
    reference_data: data_hom.zip
    color: purple
leaderboard:
  leaderboards:
    RESULTS &RESULTS
    label: Results
    rank: 1
  columns:
    coverage:
      leaderboard: *RESULTS
      label: coverage
      rank: 4
      numeric_format: 4
    precision:
      leaderboard: *RESULTS
      label: precision
      rank: 3
---
```
Competition bundle file structure

Competitions are defined by a zipped file archive (a **bundle**) containing:

- a logo image
- HTML files for the competition website
- scoring software (ZIP archive)
- gold-standard ("reference") data (ZIP archive)
- `competition.yaml`
competition.yaml settings

- Competition title, description, and logo
- Whether registration is required
- Filenames for standard and option web pages
- Configuration of competition phases
- Format of the leaderboard
competition.yaml: basic settings

title: SemEval-2017 Task 7, Subtask 3
description: Interpretation of English Puns
image: semeval2017-task7-logo.png
has_registration: True
allow_teams: True
end_date: 2017-01-30
html:
  overview: overview.html
evaluation: evaluation.html
terms: terms_and_conditions.html
data: data.html
SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

Current
- Trial
  Dec. 5, 2016, midnight UTC
- Test (Homographic)
  Jan. 23, 2017, midnight UTC

Data sets

Data sets for this subtask are described in detail on the SemEval-2017 Task 7 website.

Download
- Trial and test data

Note that, due to the difficulty in amassing a large number of pun examples per word or sense, there is no training data for this task.
Data sets for this subtask are described in detail on the SemEval-2017 Task 7 website.

Download

- Trial data
- Test data will not be released until the evaluation begins

Note that, due to the difficulty in amassing a large number of pun examples per word or per sense, there is no training data for this task.
Competition phases

- **Phases** break your competition into optional subtasks
- Each phase has its own:
  - title
  - start date
  - scoring program
  - data
  - submission limit

- Phases can be run in parallel, staggered, or (with some difficulty) in sequence
- It’s common to have a “trial” phase for sandbox testing:
  - earlier start date
  - toy data
  - unlimited submissions
competition.yaml: defining phases

phases:
  1:
    phasename: 1
    label: "Trial"
    start_date: 2016-12-05
    max_submissions: 999
    scoring_program: scorer.zip
    reference_data: data_trial.zip
    leaderboard_management_mode: hide_results
    color: white
  2:
    phasename: 2
    label: "Test\textsuperscript{(Homographic)}"
    start_date: 2017-01-23
    :
SemEval-2017 Task 7, Subtask 3
Organized by Logological - Current server time: May 7, 2017, 2:11 p.m. UTC

- **Current**
  - **Trial**
    - **Start:** Dec. 5, 2016, midnight UTC

- **Next**
  - **Test (Homographic)**
    - **Start:** Jan. 23, 2017, midnight UTC

- **Test (Heterographic)**
  - **Start:** Jan. 23, 2017, midnight

**Competition Ends**
- **Jan. 30, 2017, 11:59 p.m.**
Leaderboards

- A **leaderboard** is a dynamically updated results table
- Each phase has 0 or more leaderboards, public or hidden
- The columns are the metrics output by your scoring software
- The rows are the participants’ submissions
- You define the column labels and numeric format
- You can also rank the metrics by priority
competition.yaml: defining the leaderboards

leaderboard:
  leaderboards:
    RESULTS: &RESULTS
      label: Results
      rank: 1
  columns:
    coverage:
      leaderboard: *RESULTS
      label: coverage
      rank: 4
      numeric_format: 4
    precision:
      leaderboard: *RESULTS
      label: precision
      : ;
SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

Current			Next
Trial			Test (Homographic)

Phase description
None

Max submissions per day: 999

Max submissions total: 2

Download CSV	Download all submissions on leaderboard

<table>
<thead>
<tr>
<th>#</th>
<th>User</th>
<th>Team Name</th>
<th>F1 ▲</th>
<th>recall ▲</th>
<th>precision ▲</th>
<th>coverage ▲</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>doogyb</td>
<td></td>
<td>0.0774 (4)</td>
<td>0.0770 (4)</td>
<td>0.0778 (4)</td>
<td>0.9900 (1)</td>
</tr>
<tr>
<td>2</td>
<td>kevang</td>
<td>BuzzSaw</td>
<td>0.1544 (2)</td>
<td>0.1525 (1)</td>
<td>0.1563 (2)</td>
<td>0.9761 (2)</td>
</tr>
<tr>
<td>3</td>
<td>lhurtado</td>
<td>ELiRF-UPV</td>
<td>0.0996 (3)</td>
<td>0.0978 (3)</td>
<td>0.1014 (3)</td>
<td>0.9646 (3)</td>
</tr>
<tr>
<td>4</td>
<td>evrog</td>
<td>UTMN</td>
<td>0.0452 (5)</td>
<td>0.0424 (5)</td>
<td>0.0484 (5)</td>
<td>0.8760 (4)</td>
</tr>
<tr>
<td>5</td>
<td>tpedere</td>
<td></td>
<td>0.1557 (1)</td>
<td>0.1448 (2)</td>
<td>0.1683 (1)</td>
<td>0.8606 (5)</td>
</tr>
</tbody>
</table>
Scoring programs and reference data

- You provide a ZIP file containing arbitrary gold-standard reference data
- You provide another ZIP file containing
  - an executable **scorer**
  - a **metadata** file that describes how to run the scorer, using the following variables:
    - `$program` scorer directory
    - `$input/ref` reference data directory
    - `$input/res` submission data directory
    - `$output` output directory

description: Scoring program for Subtask 3 of SemEval-2017 Task 7 (pun interpretation)
command: java -classpath $program de.tudarmstadt.ukp.semeval2017.task7.scorer.PunScorer -i $input/ref/truth.txt $input/res/answer.txt $output/scores.txt
Scorers can be written in any language (Python, Java, Perl, etc.)

- The scorer must produce a key–value file $output/scores.txt
- Each key corresponds to a leaderboard column key:

  - coverage: 1.000
  - precision: 0.825
  - recall: 0.775
  - f1: 0.799

- stderr is captured and reported to the submitter
- stderr and stdout are captured and stored for the competition organizer
Data sets for this subtask are described in detail on the SemEval-2017 Task 7 website.

Test data will not be released until the evaluation begins.

Note that, due to the difficulty in amassing a large number of pun examples per word or per sense, there is no training data for this task.
To create a competition, upload a bundle with the definition of the new competition.

Upload definition...
Upon uploading a competition

- You will (usually) be warned if there is a problem
- You optionally “publish” your competition:
  - Unpublished: accessible only if you know the URL (default)
  - Published: listed on CodaLab Competitions home page
- Competitions can be edited via the web interface to:
  - Publish/unpublish
  - Change competition settings
  - Edit web pages (via rich text editor)
  - Add or remove reference data and scoring programs
- Editing the original YAML is no longer possible!
Edit Competition

General Information

Title: SemEval-2017 Task 7, Subtask 3

Description: Interpretation of English Puns

- [ ] Disallow leaderboard modifying
- [ ] Force submission to leaderboard

Logo: Currently: logos/semeval2017-task7-logo_33.png  [Clear]

Change: [Browse...] No file selected

- [x] Registration Required

End Date (UTC): 2017-01-30 23:59:59

- [x] Publicly Available
- [ ] Enable medical image viewer
- [ ] Enable detailed results
<table>
<thead>
<tr>
<th>NAME</th>
<th>KEY</th>
<th>DESCRIPTION</th>
<th>TYPE</th>
<th>ACTIONS</th>
<th>MULTI-DELETE?</th>
</tr>
</thead>
<tbody>
<tr>
<td>scorer_1_15705</td>
<td>66628265-3f45-4a9f-bb08-ca21ed02b33b</td>
<td>None</td>
<td>Scoring Program</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>data_trial_1_15705</td>
<td>ced44e66-4c5c-4a40-86ad-e35bc6385d4b</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>scorer_1_15706</td>
<td>fc3a8853-0490-47d8-b3f4-9cb27e40b4</td>
<td>None</td>
<td>Scoring Program</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>data_trial_1_15706</td>
<td>d81e2ebf-e875-431d-8c16-281670513146</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>scorer_1_15707</td>
<td>96adea9b-0957-4e00-84ff-8cdd01ec5fd</td>
<td>None</td>
<td>Scoring Program</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>data_trial_1_15707</td>
<td>6159a63c-e180-40b8-9ec7-6594a0482b97</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>scoring_program_1_15847</td>
<td>ae8efe2a-d979-4444-bcf3-7316b9d6ba9d</td>
<td>None</td>
<td>Scoring Program</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>dev_data_1_15847</td>
<td>d8277bc8-3660-4329-ad47-f09f610c5bb</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>test_data_2_15847</td>
<td>a666e75d-869d-4dde-bfe1-4fb727c85d6e</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>SemEval 2017 Task 7</td>
<td>b6b087f4d-02c1-4682-b1ce-33894c1b5002</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>Subtask 1 Homographic</td>
<td>ebf0c1b99-cf5f-4471-6af8-adf955d62073</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
<tr>
<td>SemEval 2017 Task 7</td>
<td>66628265-3f45-4a9f-bb08-ca21ed02b33b</td>
<td>None</td>
<td>Reference Data</td>
<td>Download</td>
<td></td>
</tr>
</tbody>
</table>
The “Submissions” page allows you to:

▶ see a list of all successful and failed submissions
▶ download the original system output
▶ download a CSV of all system scores
▶ re-run the scorer on one or all submissions
▶ delete a submission
▶ (un)hide a submission from the leaderboard
### Submissions

**Phase**

Test (Homographic) - Jan. 23, 2017, midnight UTC

**View all submission's metadata**

**Select phase description**

None

**Max submissions per day:** 999

**Max submissions total:** 2

**Submissions**

<table>
<thead>
<tr>
<th>#</th>
<th>SUBMITTED BY</th>
<th>SUBMITTED BY</th>
<th>SUBMISSION ID</th>
<th>SUBMISSION ID</th>
<th>FILENAME</th>
<th>STATUS</th>
<th>LEADERBOARD</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kevang</td>
<td>Jan. 25, 2017, 5:41 p.m.</td>
<td>385636</td>
<td>WSDse_split.zip</td>
<td>Finished</td>
<td>True</td>
<td>0.9761</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>evrog</td>
<td>Jan. 27, 2017, 3:34 a.m.</td>
<td>385821</td>
<td>homes_3.zip</td>
<td>Finished</td>
<td>True</td>
<td>0.8760</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>lhurtado</td>
<td>Jan. 27, 2017, 7 p.m.</td>
<td>385895</td>
<td>answer.bt_task7.zip</td>
<td>Failed</td>
<td>False</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>lhurtado</td>
<td>Jan. 27, 2017, 7:10 p.m.</td>
<td>385897</td>
<td>answer.bt_task7.zip</td>
<td>Finished</td>
<td>True</td>
<td>0.9646</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>evrog</td>
<td>Jan. 28, 2017, 4:02 a.m.</td>
<td>385931</td>
<td>homes_3_extended.zip</td>
<td>Failed</td>
<td>False</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>evrog</td>
<td>Jan. 28, 2017, 4:20 a.m.</td>
<td>385932</td>
<td>homes_3_extended.zip</td>
<td>Finished</td>
<td>False</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>
Student competitions
**Mini Challenges 2017**

<table>
<thead>
<tr>
<th>Challenges (M2 students)</th>
<th>Abstract</th>
<th>Task</th>
<th>Solution (Video L2 students)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blue</strong></td>
<td><strong>Activity of molecules against HIV</strong></td>
<td>The objective is to predict which compounds are active against the AIDS HIV infection. The dataset has two classes: active or inactive (Binary Classification). The variables represent properties of the molecule inferred from its structure. Note: this project is running on the LRI server. In case of problem, a previous version on the main Codalab instance is available.</td>
<td><strong>Marine</strong></td>
</tr>
<tr>
<td></td>
<td>The problem is to relate molecular structure to activity to screen new compounds before actually testing them with High Throughput Screening (HTS) in vitro experiments. HTS is a method for massive scientific experimentation used in drug discovery, linking the fields of biology and chemistry. This method remains very costly process despite many recent technological advances in the field of biotechnology. This is why applying machine learning methods would be of great benefit for the pharmaceutical industry to reduce the number of compounds that need to be tested.</td>
<td></td>
<td><strong>Cobalt</strong></td>
</tr>
<tr>
<td><strong>Ocyan</strong></td>
<td><strong>Lothlorien</strong></td>
<td>A computer vision challenge is proposed for undergraduate students in which the challenger must predict the class of a person (major or minor) based on a picture of his/her face. Note: the main Codalab instance of this challenge has been tested. Note: this project is running on the LRI server. In case of problem,</td>
<td><strong>Cerulean</strong></td>
</tr>
<tr>
<td></td>
<td>This challenge aims at addressing the issue of resources access (website, drug purchase, violent movie, etc.) based on the age of a person. Indeed a lot of violent content is accessible on the internet and 45% of children under 12 are not monitored by parental control. For this sake, we rely on the person's</td>
<td></td>
<td><strong>Turquoise</strong></td>
</tr>
</tbody>
</table>
Student competitions

- skill development for student participants/organizers
  - experimentation, problem formulation/solving
  - writing/fulfilling technical specifications
  - collaboration with peers
  - presentation of research results

- benefits for student participants
  - unrestricted choice of tools/methods
  - instant feedback to students
  - freedom to experiment

- benefits for teachers
  - automatic enforcement of submission deadlines
  - instant tabulation of scores
  - consistent packaging of submissions
Student competitions

- skill development for student participants/organizers
  - experimentation, problem formulation/solving
  - writing/fulfilling technical specifications
  - collaboration with peers
  - presentation of research results

- benefits for student participants
  - unrestricted choice of tools/methods
  - instant feedback to students
  - freedom to experiment
Student competitions

- skill development for student participants/organizers
  - experimentation, problem formulation/solving
  - writing/fulfilling technical specifications
  - collaboration with peers
  - presentation of research results

- benefits for student participants
  - unrestricted choice of tools/methods
  - instant feedback to students
  - freedom to experiment

- benefits for teachers
  - automatic enforcement of submission deadlines
  - instant tabulation of scores
  - consistent packaging of submissions
Caveats
Caveats

CodaLab Competitions is . . .

▶ undocumented
▶ unpredictable
▶ unintuitive
▶ unstable

But it's . . .

▶ free (as in beer)
▶ free (as in speech)
▶ popular
▶ under active development
▶ probably a time-saver, in the long run
Caveats

CodaLab Competitions is . . .

▶ undocumented
▶ unpredictable
▶ unintuitive
▶ unstable

But it’s . . .

▶ free (as in beer)
▶ free (as in speech)
▶ popular
▶ under active development
▶ probably a time-saver, in the long run
Thank you!

Questions?