i2b2

Informatics for Integrating Biology & the Bedside

A National Center for Biomedical Computing

About Us   Software   NLP Data Sets   i2b2 tranSMART Foundation

UNIVERSITY AT ALBANY
State University of New York

2008 NLP Shared Task

- Home
- Workshop Schedule
- Challenge Results
- Challenge Material Repository
- *Administration*
  - Registration
  - Download Tracking
  - Uploaded Files
- *Data*
  - Download
  - *Upload*
    - Obesity
- *User*
  - Profile
- Documentation
- FAQ
- About
- Current Challenge
- NLP Research Data Sets

Documentation

i2b2 Obesity Challenge Documentation

Overview: The i2b2 Obesity Challenge asks you to build automatic systems that will identify obese patients and the co-morbidities exhibited by them based on the narrative patient record. The focus of the challenge is obesity and 15 of its best represented co-morbidities in the Research Patient Data Repository (RPDR) of Partners HealthCare.

The Data: The data for the challenge were randomly drawn from the RPDR using a query that drew records of patients who were evaluated for either obesity or diabetes. Each of the records in the data includes zero to more than ten occurrences of the stem "obes". The drawn records were de-identified semi-automatically. An automatic pass, followed by two parallel manual passes were made over each record. These were followed by a third manual pass that resolved the disagreements between the two parallel manual passes. In order to make the data HIPAA compliant, patient names, health proxy and patient family member names, doctor names, hospital names, ID numbers, phone and pager numbers, dates, locations, ages, mentions of companies related to patient's occupations, some nationalities, and other potential identifiers were replaced with surrogates. The surrogate replacement process inserted random names from the US Census Bureau Database for each of the patient, health proxy, family member, and doctor names in the data. We made no effort to keep co-reference, though any name could have been drawn from the US Census Bureau Database more than once. For hospital names, ID numbers, phone and pager numbers, dates, locations, and ages, we generated surrogates randomly. Mentions of companies or workplaces were treated like locations. Nationalities were replaced with other random nationalities.

The challenge data were annotated by two obesity experts from the Massachusetts General Hospital (MGH) Weight Center. The two experts summarized the obesity and co-morbidity information in each patient record. For obesity and for each co-morbidity, they provided document-level judgments that marked whether the disease/co-morbidity was present (marked with a "Y" that stands for "Yes, the patient has the co-morbidity"), absent (marked with a "N" that stands for "No, the patient does not have the co-morbidity"), questionable (marked with a "Q" that stands for "Questionable whether the patient has the co-morbidity"), or unmentioned (marked with "U" that stands for "the co-morbidity is not mentioned in the record"). They provided judgments that were strictly based on text; we refer to these judgments as "textual judgments" and mark the source of these judgments as "textual" in the annotation files. The disagreements among the obesity experts on their textual judgments were resolved by a resident doctor at the MGH. The majority vote among the three doctors determined the final textual judgments. The two obesity experts also provided their "intuitive judgments" on obesity and co-morbidities. Intuitive judgments are based on implicit information in the narrative text. We mark the source of these judgments as "intuitive" in the annotation files. No tie-breaker was available for resolving the disagreement on intuitive judgments; therefore, only the records that the two experts agreed on have been released. Possible intuitive judgments are limited to "Y", "N", and "Q" because "U" is irrelevant as an intuitive judgment. The Kappa agreement between the annotators (before any tie-breaking) on each of the co-morbidities are as follows:

Co-morbidity	Textual Kappa	Intuitive Kappa
Obesity	0.9068	0.8617
Diabetes mellitus (DM)	0.9072	0.8685
Hypercholesterolemia	0.8732	0.6772
Hypertriglyceridemia	0.7123	0.7154
Hypertension (HTN)	0.8150	0.6719
Atherosclerotic CV disease (CAD)	0.7790	0.8140
Heart failure (CHF)	0.8122	0.7394
Peripheral vascular disease (PVD)	0.9423	0.7328
Venous insufficiency	0.7882	0.4369
Osteoarthritis (OA)	0.7551	0.7587
Obstructive sleep apnea (OSA)	0.9193	0.9178
Asthma	0.9041	0.7638
GERD	0.8856	0.5865
Gallstones / Cholecystectomy	0.9198	0.8987
Depression	0.9170	0.8613
Gout	0.9318	0.9218

The training set includes 60% of each of "Y", "N", "Q", and "U" textual judgments for obesity from the complete data set. For the thus drawn records, we also provide intuitive judgments. No special effort is made to provide a 60%-40% split between training and test sets on the co-morbidities. The frequencies you observe of various co-morbidities and various judgments on them reflect the true frequencies of these co-morbidities in the Partners HealthCare RPDR.

The Challenge: The participants are asked to build systems that will correctly replicate the textual and intuitive judgments of the obesity experts on obesity and co-morbidities based on the narrative patient records. Participants may choose to focus on only predicting the textual judgments, on only predicting the intuitive judgments, or on both. The efforts of the participants will be evaluated on the test data, to be released on June 23, 2008 for only three days. Development on the test set is

not allowed. We ask that the participants return to us, through the upload page, the output of their system runs on the test data. Each team is allowed up to three system runs for predicting textual judgments and up to three system runs for predicting the intuitive judgments on the test data. The predicted textual judgments and the predicted intuitive judgments will be scored independently of each other. Given the difference in the sets of records that are relevant to each co-morbidity and the difference in the sets of records that are relevant to textual and intuitive judgments, the participants will be provided with template xml forms on which they can provide their predictions for the test set. The template xml forms will be released along with the test records.

Participants are allowed to use existing systems and tools for the obesity challenge. We also allow the use of external data sources, such as extra patient records and extra annotation efforts, to complement our data. However, we ask that teams that have access to such rare and invaluable sources declare this information on the submission page.

Competition: The i2b2 Obesity challenge is conducted as a competition. The system outputs provided to i2b2 will be evaluated and ranked based on their performance on the test data.

Evaluation metrics: Precision, recall, and F-measure will be used as evaluation metrics. Given the non-uniform distribution of records into various classes, macro-averaged F-measure will be the primary and micro-averaged F-measure will be the secondary metric for ranking. You can download the evaluation scripts that implement these metrics in the form of a .jar file or obtain the source code. If you choose to use the .jar file, you will need Jdk1.5 or higher. You may invoke the evaluation script with the command 'java -jar evaluation.jar <ground_truth> <system_output>' where ground_truth is the correct annotations (e.g., training file) in xml format and system_output is the predictions provided by your system, in the same format as the ground_truth file. The evaluation code displays the results on the screen.

Workshop: A subset of the challenge participants will be invited to present their work at a day-long workshop organized in conjunction with AMIA 2008. i2b2 will also work with a journal to arrange for a special issue focused on the obesity challenge and the participating systems.

[back to top]

Home l Contact l Sitemap l Search

PARTNERS.   ©2005 - 2019
                   Partners Healthcare

design by Digizyme