

Lecture 14: Recent advances & industry

STATS 202: Data Mining and Analysis

Linh Tran

tranlm@stanford.edu



Department of Statistics
Stanford University

August 11, 2021



- ▶ HW4 due due in 9 days.
- ▶ Final project submissions close in 12 days.
 - ▶ Write-up due via Gradescope due in 16 days.
 - ▶ Reference your Kaggle leaderboard name.
- ▶ HW3 almost done grading.



- ▶ Advances in NLP
- ▶ Industry applications



Project #1: Classification.

Stanford STATS 202 Classification 2020

Predict flagged assessments

26 teams · a day to go

[Overview](#) [Data](#) [Notebooks](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [Host](#) [My Submissions](#) [Submit Predictions](#)

Overview Edit

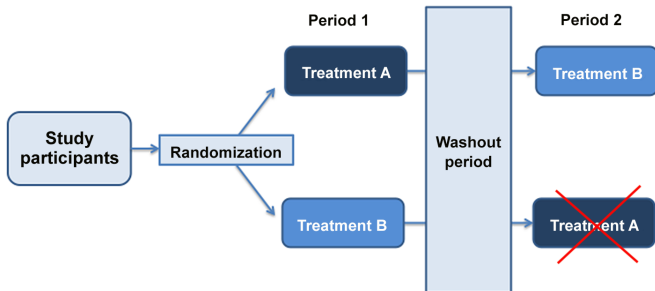
Description

Evaluation

Timeline

Having humans audit all of the PANSS assessments can be time consuming and expensive. It is therefore reasonable to wonder if a machine learning algorithm can reliably predict which of the assessments are erroneous. Having a reliable algorithm would allow the clinical auditor to focus only on the subset of assessments that have issues, instead of all of the assessments uniformly. Therefore, your goal is to predict which of the assessments in Study E will be either flagged for review or assigned to a CS.

Project #2: Crossover effects.



Project #3: Record linkage.

Hospital A

Name	Sex	SSN	Age	Height (cm)
Angel Smith	Male	002-98-3445	20	180
Divine Scavo	Female	001-34-2356	24	162.5
Selene Paul	Female	000-22-6509	22	160
Sandrine Pal	Female	009-12-2222	23	167.5

Dataset



Hospital B

Name	Sex	SSN	Age	Height (cm)
Angel Smith	Male	002-98-3445	20	180
Divine Scavo	Female	001-34-2356	24	162.6
Ryan Solis	Male	033-24-0281	18	157.5
Katie Gomes	Female	243-30-2470	20	175

Dataset



Project #4: *Extracting Symptoms and their Status from Clinical Conversations.*

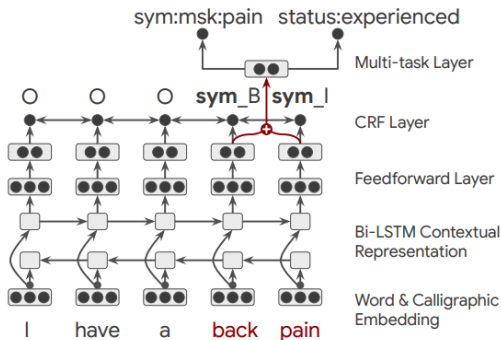


Figure 1: The architecture of Span-Attribute Tagging (SA-T) Model, illustrating the span extraction layer followed by the attribute tagging layer.



- [1] Ruder, Sebastian et al (2018). *Frontiers of Natural Language Processing*.