Lecture 14: Recent advances & industry

STATS 202: Data Mining and Analysis

Linh Tran

tranlm@stanford.edu

Department of Statistics
Stanford University

August 11, 2021

Announcements

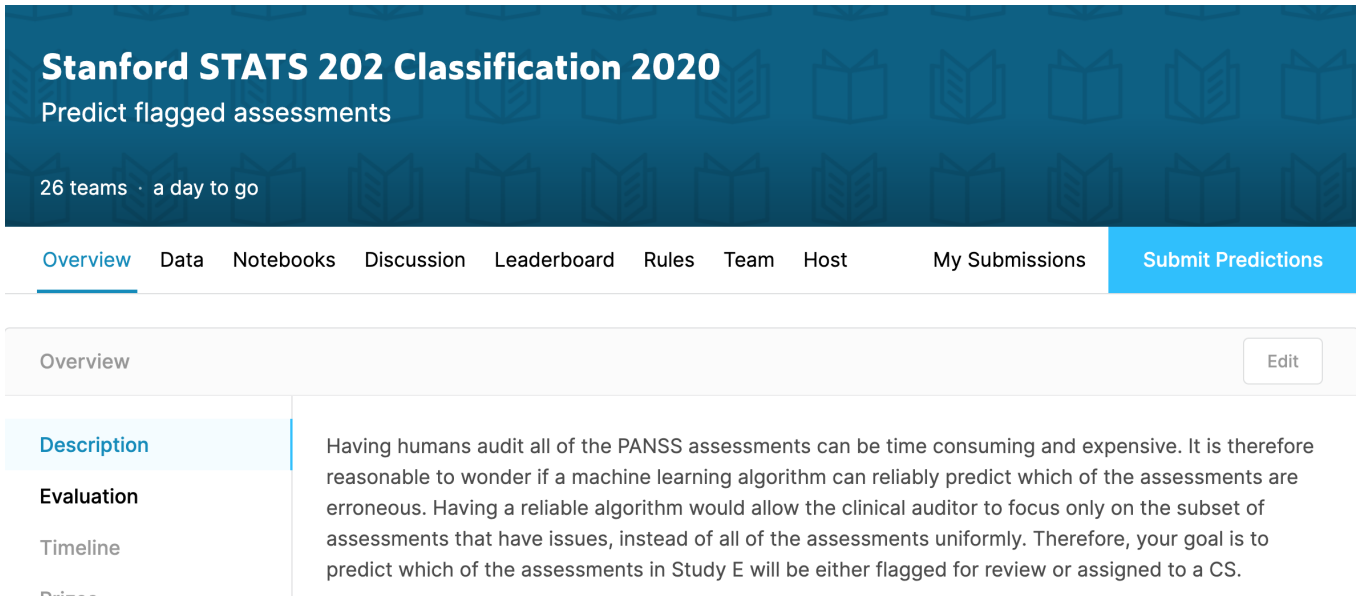
- HW4 due due in 9 days.
- Final project submissions close in 12 days.
  - Write-up due via Gradescope due in 16 days.
  - Reference your Kaggle leaderboard name.
- HW3 almost done grading.

Outline

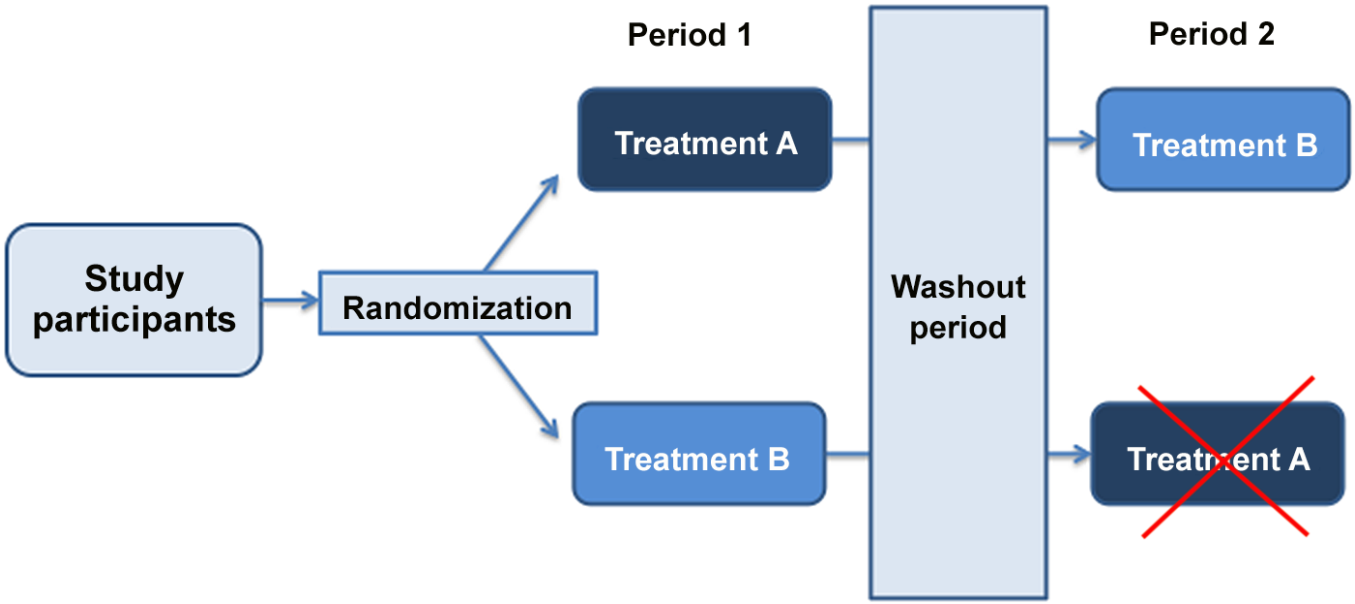
- Advances in NLP
- Industry applications

Projects

Project #1: Classification.



Project #2: Crossover effects.



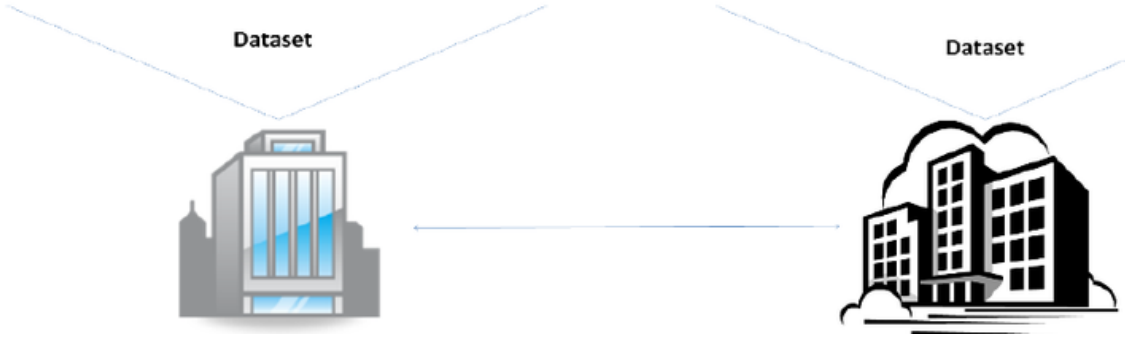
Project #3: Record linkage.

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

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



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

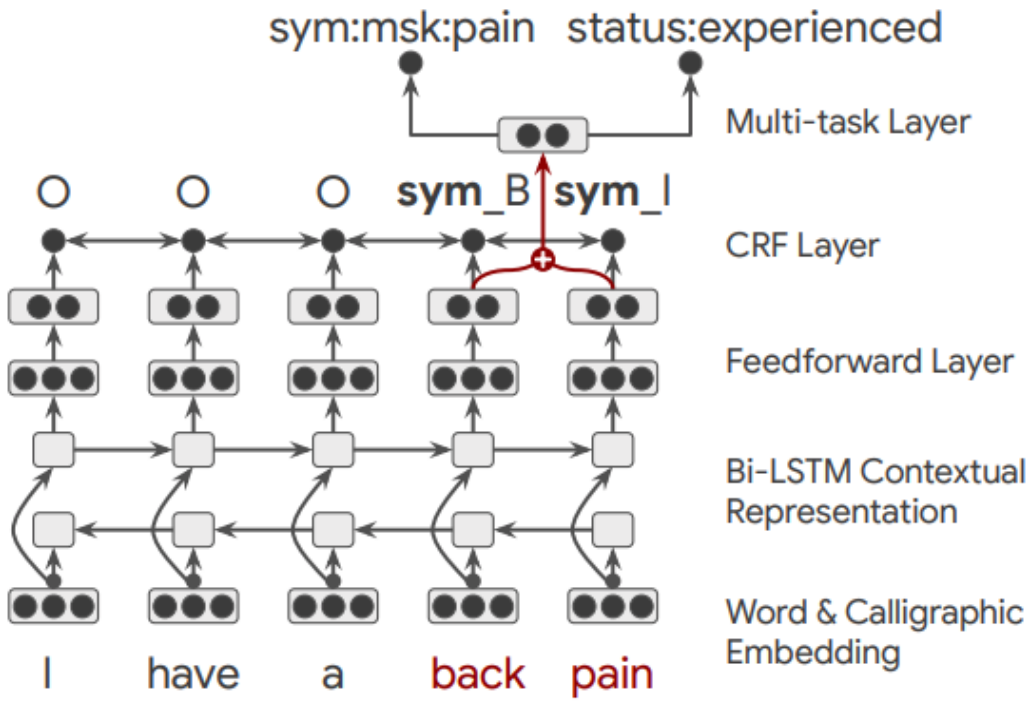


Figure 1: The architecture of Span-Attribute Tagging (SA-T) Model, illustrating the span extraction layer followed by the attribute tagging layer.

References

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