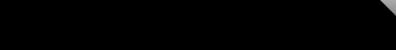


The New York Times

Crossword with NLP

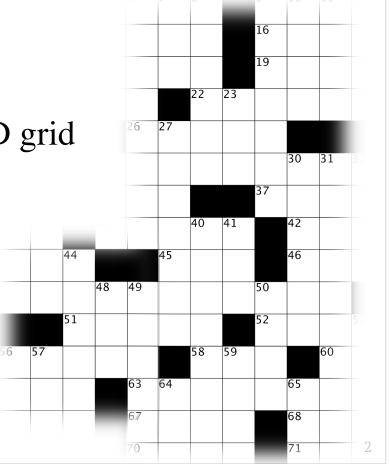


- Tony Ghabour
- February 28, 2020





- Popular daily word puzzle on 2D grid
- Grid typically 15 x 15
- Blocks define boundaries,
 specify answer lengths
- Occasionally have a theme



T H E P U Z L E

- Short clues used to fill grid
- Multi-word answers permitted
- Solutions are unique
- Difficulty increases as week progresses

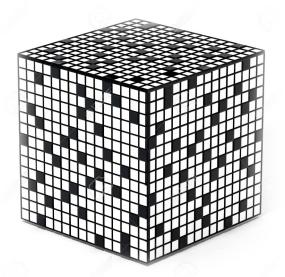
T H E Q U E S T I O N

The **Sunday** puzzle is famously **harder**, even among enthusiasts.

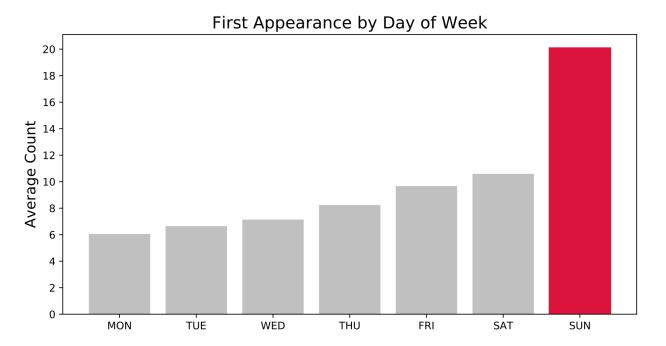
Why?

T H E C O R P U S

- Over 14 thousand puzzles
- Nearly 1.3 million clues and answers
- Dates range from 1977 to 2017
- Spans 1716 authors and two editors



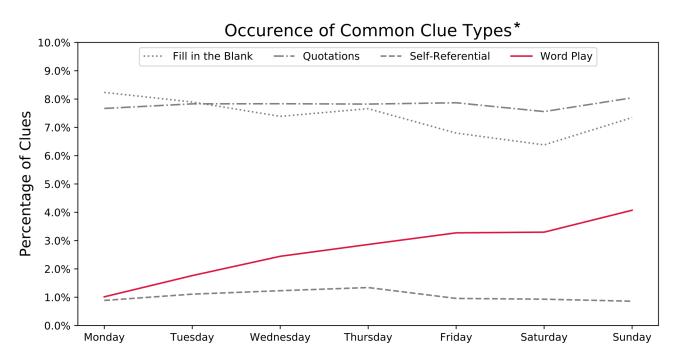




Sunday includes twice as many novel answers compared to next closest day.

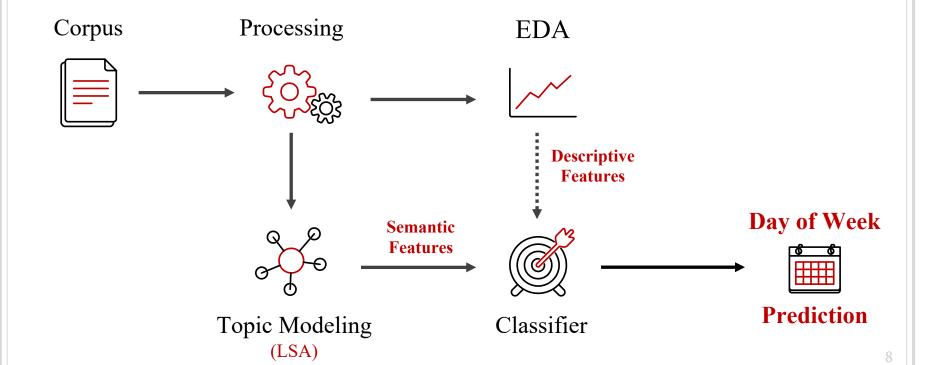
E

* Descriptive Features



Prevalence of word play increases throughout the week...contributing to difficulty.





T | **O** | **P** | **I** | **C** | **M** | **O** | **D** | **E** | **L** | **I** | **N** | **G**



Self-referential: starred clue, hint, answer, puzzle



Films: Oscar winning film/director/role/actress



Star wars: villain, droid, princess, character, Solo



International: French coin, Hebrew letter, German city



U.S. Cities: Phoenix, Dayton, San Francisco, Provo

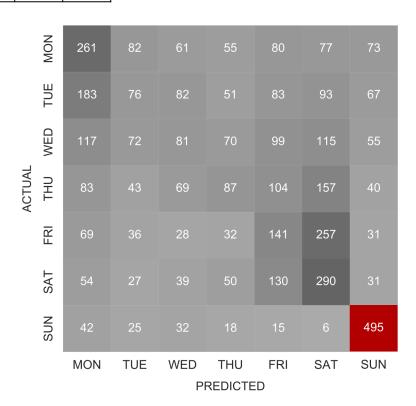


Sports: baseball, basketball, football, NFL, hall of fame

T H E R E S U L T S

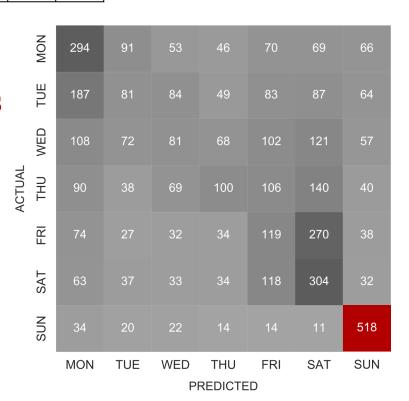
- Sematic features alone

 (i.e. topics) enough to
 specify Sunday puzzles
- Saturday and Monday puzzles also readily identified



 $T \mid H \mid E \mid B \mid R \mid E \mid S \mid U \mid L \mid T \mid S$

- Adding descriptive features generally improves classification for all days
- However mid-week puzzles remain a challenge





- Sundays feature the most first-time answers
- Also reflect highest rate of word play
- Readily identifiable based on topic modeling
- Mid-week puzzles more difficult to classify (puzzle difficulty is subjective...or are the editors wrong?)

Thank You

Questions?

