

Validation and Using Innovations in NLP to Study Innovation

Ina Ganguli (UMass Amherst)

Jeffrey Lin (Phil FRB)

Vitaly Meursault (Phil FRB)

Nick Reynolds (Essex)

NBER I3 Meetings

December 3, 2022



How can we measure similarity between patents?

- Many cases where we want a measure of how *similar* patents are to other patents (to identify novelty, to identify 'control' patents, etc.)
- Traditional approaches: Patent Classification systems
- Growing use of text analysis/NLP to study innovation, including measures of patent similarity (e.g. Feng, 2020, Arts et al. 2018, papers from yesterday...)

Innovations in NLP



... but should innovation economists care?

- Maybe the improvements are marginal
- Maybe the improvements don't translate to innovation contexts
- Maybe we should just prefer simpler models
- Can we even tell which one is better?
- ***What does it mean for similarity measures to be better?***

Validation to understand which model is better for a specific case

1. Interference cases

- *Legally meaningful similarity*
- Modern patents & applications (2001-2011)

2. Human validation task

- *Human sense of similarity*
- Historical patents (1880-1920)

(1) Patent Interferences

*Interfering patent application contains claims with the same (or substantially the same) subject matter as a previous patent i.e. **patent interference cases are indicators of very high level of similarity***

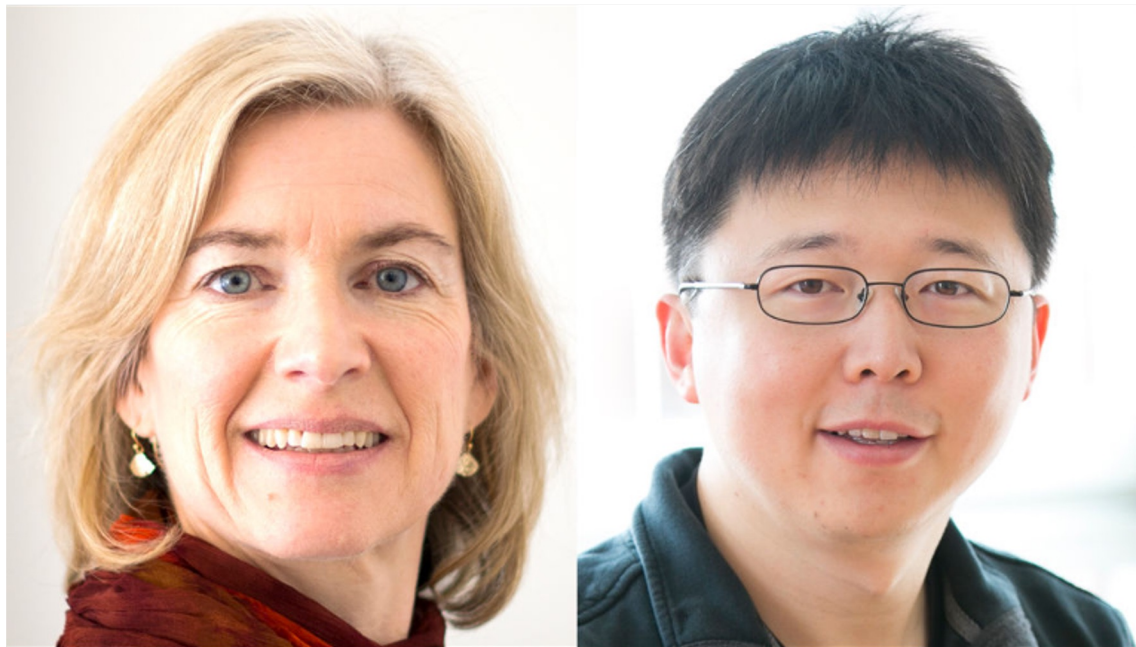
INTERFERENCES.

92. An interference is a proceeding instituted for the purpose of determining the question of priority of invention between two or more parties claiming substantially the same patentable invention. The fact that one of the parties has already obtained a patent will not prevent an interference; for, although the Commissioner has no power to cancel a patent, he may grant a patent for the same invention to another person who proves to be the prior inventor.

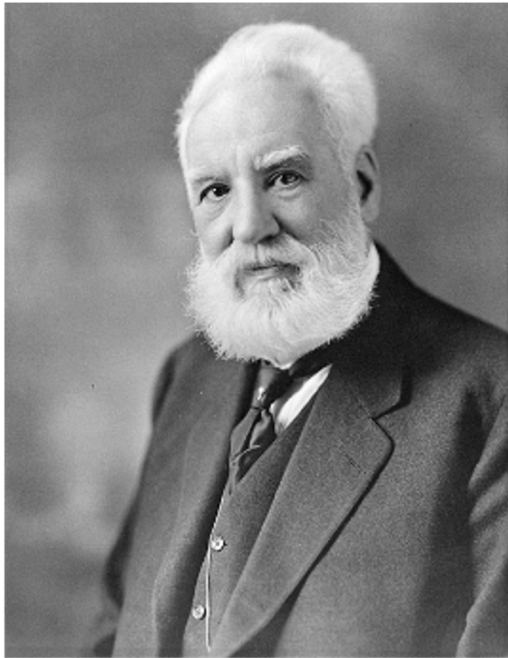
Rev. Stat. sec.
4904.
1878—51, 52.
Interference de-
fined.

Rev. Stat. sec.
4904.

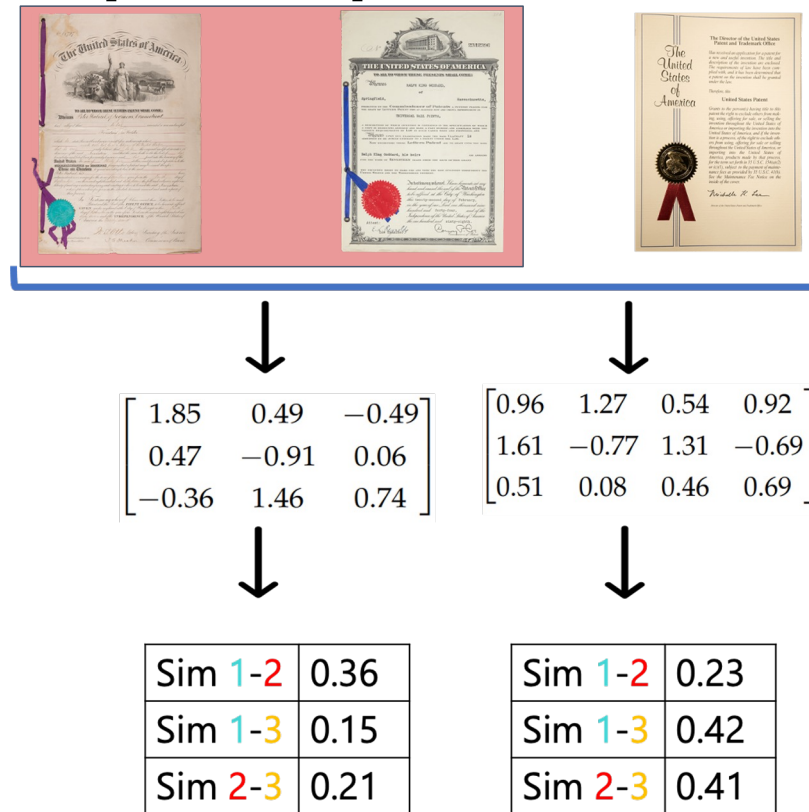
Doudna & Zhang (CRISPR, December 21, 2015)



Graham Bell & Gray (Telephone, February 14, 1876)



Interference as patent-pair classification task

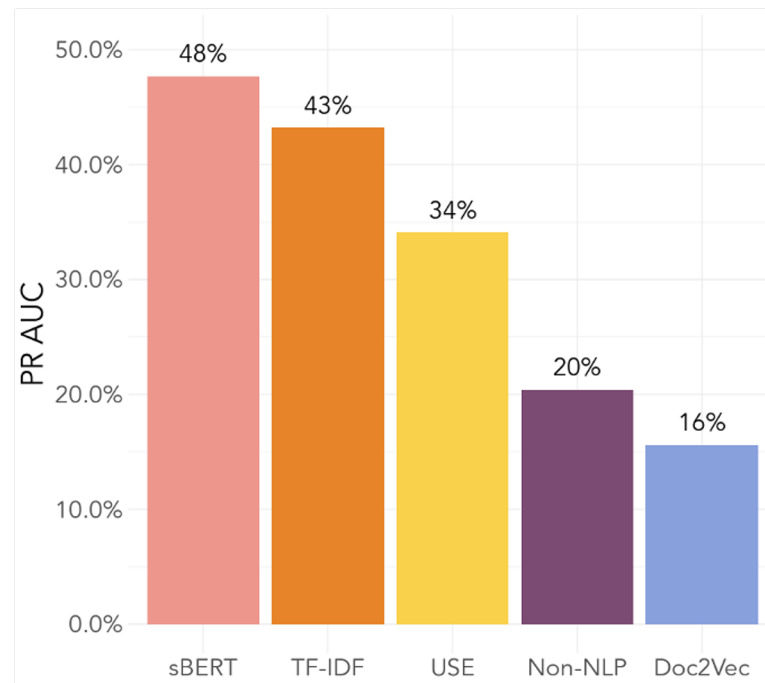


More complex isn't always better for interference classification

Sentence-BERT
performs best

But TF-IDF performs
second best

USE and Doc2Vec
don't work that well



(2) Historical patent similarity requires human validation

- Data generating process for historic patents is different - can't guarantee that the best representation is the same as for modern patents
- Can't use external measure like interference cases (since no patent applications)
- But we can validate model outputs against human judgement
- We focus on comparing to models performing best with interferences – sBERT and TF-IDF
- **Our tasks is designed to tell us which model does better**

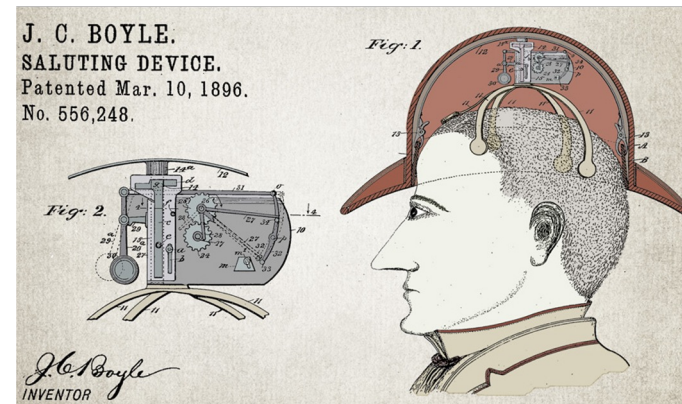
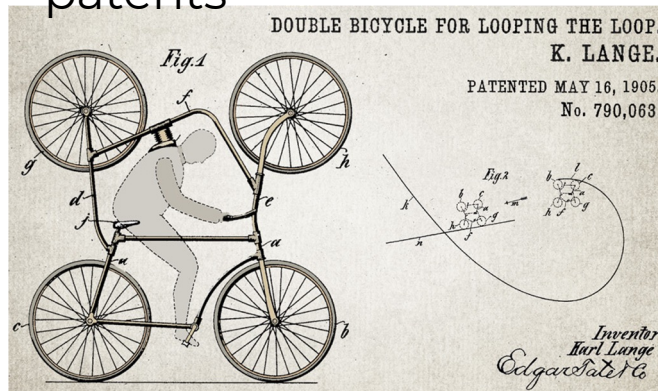
Setting up a validation task is not straightforward...

Straightforward idea:

How similar is this pair of patents?

Problems:

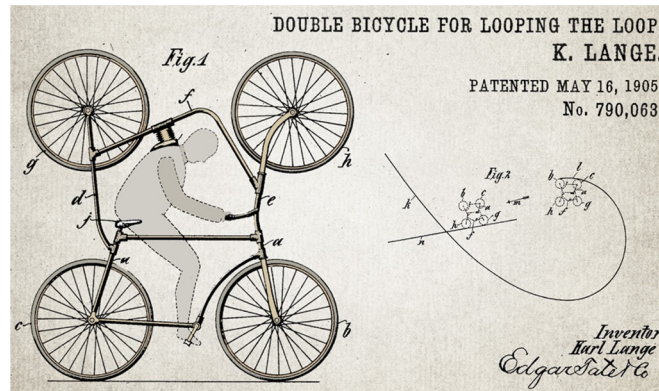
- Degree of similarity is hard to quantify
- Unclear what to do with very similar or very different patents



We elicit human judgement with informative triples

Is Focal Patent more similar to Patent 1 or 2?

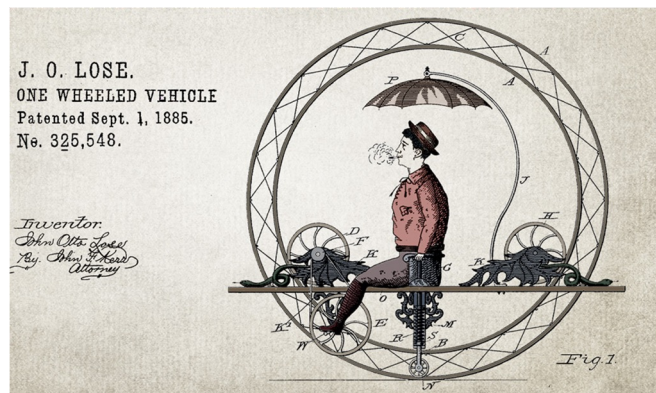
Focal Patent



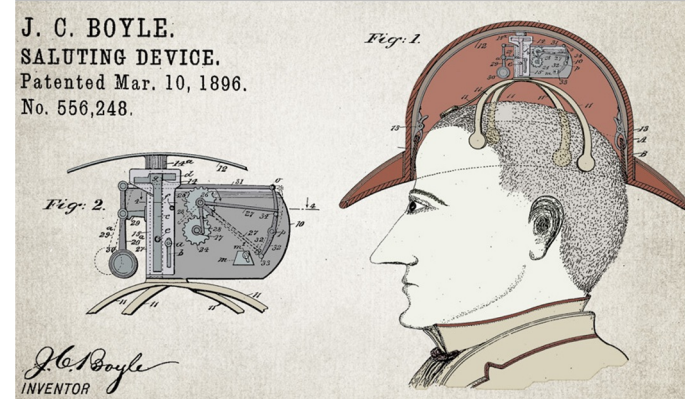
Triple sampling:

- 1) Models disagree
- 2) Each model thinks all three patents are somewhat similar
- 3) Each model thinks most similar pair is extremely similar

Patent 1



Patent 2



Actual example

Making it easy for humans:

Extract “improvements in”
part and part of the claims

Choices: 1, 2 or 0 (can’t
answer)

**Humans break ties in
sBERT's favor:**

4 coauthors: 89% of the time

5 RA's: 75% of the time

text_focal	text_1	text_2
IMPR: improvements fully set forth in the following specification	IMPR: Improvement in Pneumatic Tires	IMPR: Improvements in Wheels
CLAIMS: 1. In a wheel for vehicles, the combination of the hub with the T-shaped tire, the spokes, and the shoes to receive the outer	CLAIMS: A pneumatic tire consisting of an air-tube, a detachable cover consisting of a flexible material, said cover having at each e	CLAIMS: 1. In a wheel, a felly having grooves formed in its inner face around the spoke-holes, or partially around the spoke-holes, i

Cautious optimism about NLP innovations

We show that for similarity measures, the best performing model is sentence-BERT, outperforms TF-IDF in our main validation exercises

- The improvement is economically large
- Sentence BERT has large compute times without GPU's but is otherwise very simple to use

But:

- Other models considered (Doc2vec and universal sentence encoder) perform worse than TF-IDF
- So a more complex model isn't always better

Bottom line: validate, validate, validate!