Predicting Patentability: An AI Approach to Identifying the Dominant Future Technologies

Tara Sowrirajan

In collaboration with

Sourav Medya

Ryan Whalen

Brian Uzzi

Hit Technology Detection

- Tech & patents are strategic assets
  - Boosts economic growth
  - Raises firm value: (Trajtenberg (2005): Citations per patent increase firm value by 3 percent, r=0.76)
  - Investments - telephone patent was worth 25 million to Bell
  - Turns science into life changing realities

Future technology is challenging to predict!

Wartburg et al. 2005

Sources: (Carpenter et al,. Citation rates to technologically important patents. World Patent Information. 1981); Narin et al., 1987; Trajtenberg, 1990; Griliches, 1990; Deng et al., 1999; Breitzman and Thomas, 2002; Harhoff et al., 2003).

Can AI discover the innovation frontier?

- Data: patent applications from 3 countries
- AI model to predict hit technologies from free text
- Interpretability: AI finds hidden, interdependent clusters of dominant future technology

Research Design

Can we make a model that mimics human decisions on what get patented *before* we predict what becomes hit patents of the future?

Predict hit inventions

- Can we save resources by assessing how good an application is?

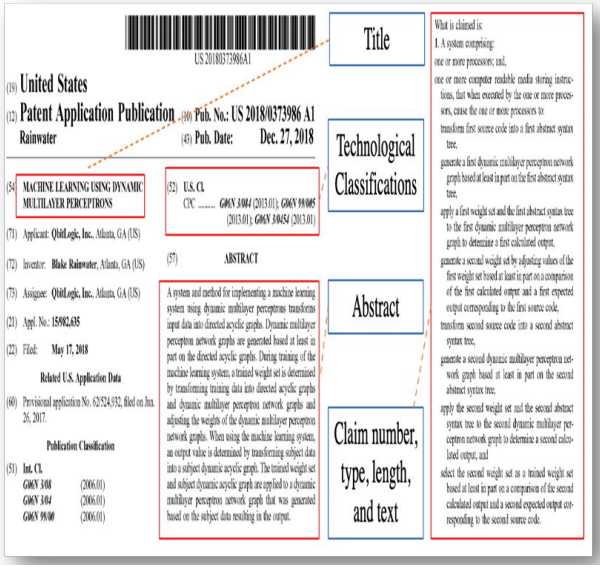
- What is hit innovation?
  - Patent impact through citations
- Large corpus of data to solve this problem

Generic AI framework to do both of these?

Data Extraction

3 International Patent Datasets

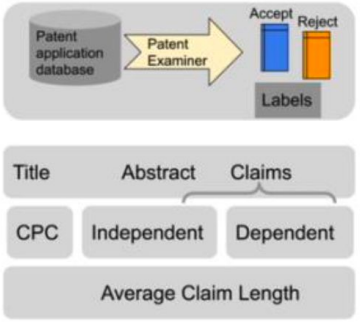
- 4,000,000 USA
- 1,000,000 UK
- 500,000 Canada



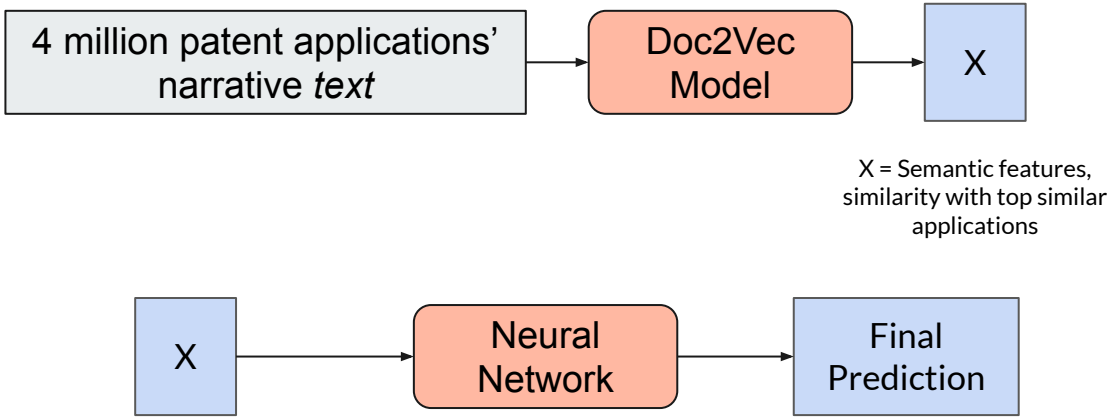
Feature Extraction

- Free text (application)
  - Abstract
  - Claims
  - Narrative approach
- Numerical features
  - # inventors
  - CPC technical classification
  - # capability claims
  - Claim length
  - Inventor part of company
  - Metadata approach

Data Overview

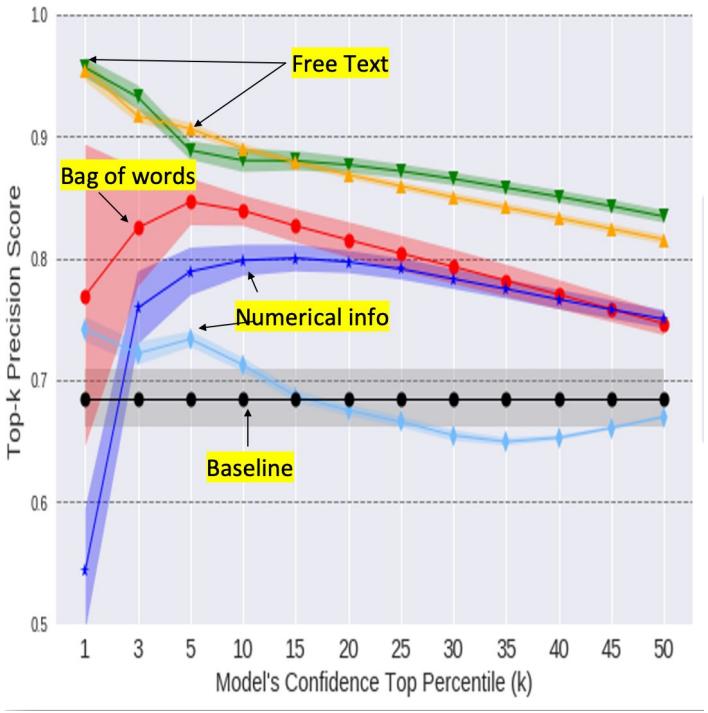


Pipeline



Data Set	Train Range	Train Size	Test Range	Test Size
#1	2001 to 2007	1.1 M	2008 to 2009	464K
#2	2010 to 2016	2.2 M	2017 to 2018	191K
#3	2001 to 2015	3.5 M	2016 to 2018	449K

Accuracy of AI Model

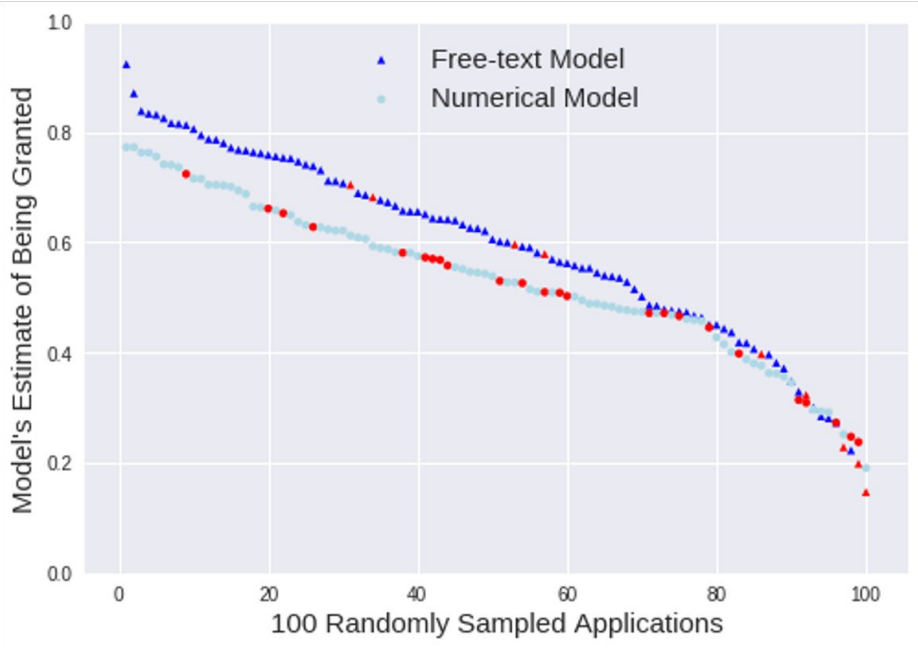


Top K Performance
4 Comparative Models

(Baseline, Numerical Models, Bag of Words, Free Text)

Free text of patent predicts better than numerical info!

Accuracy of AI Model



Predicted probabilities of being granted for randomly selected applications

Free text model rarely makes any mistakes, and above 70% confidence is 100% accurate!

Tests on Ground Truth Decisions

How can we test baked in variance of examiners decisions?

Ground Truth Appeals Data Test

Independent board of examiners "retry" original decision?

AI avoids 47% of wrongful denials

AI confirms 69% of rightful rejections

Predicting Hit Technologies

Our model is 95 % accurate for top 30% confident predictions

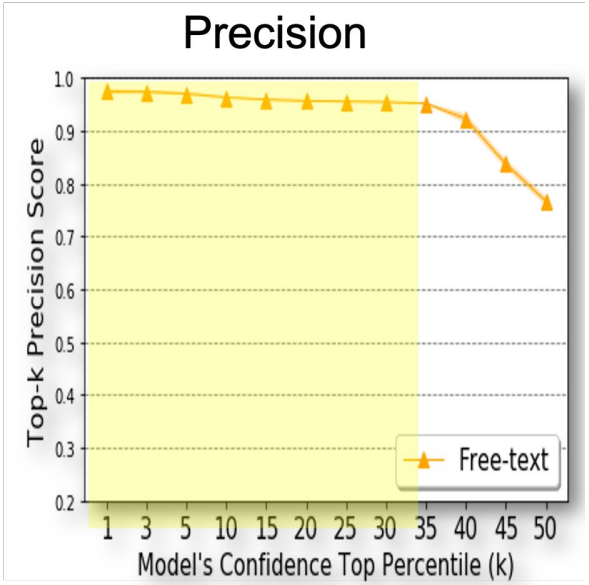
Predict Hit vs Flop Invention

Retrain NN Model

- Hit inventions are top 10% citations
- 8 years of citations

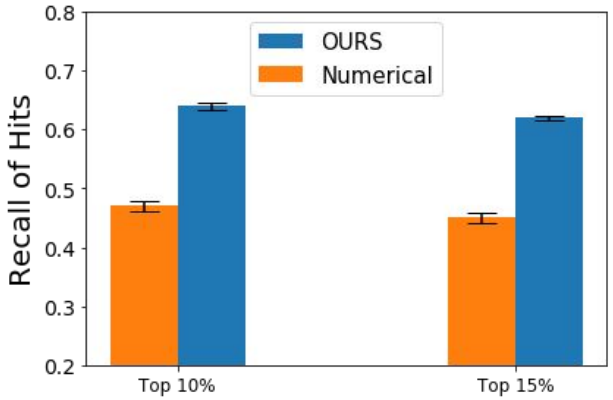
Training data: 2001 - 2009

Testing data: 2010 - 2011



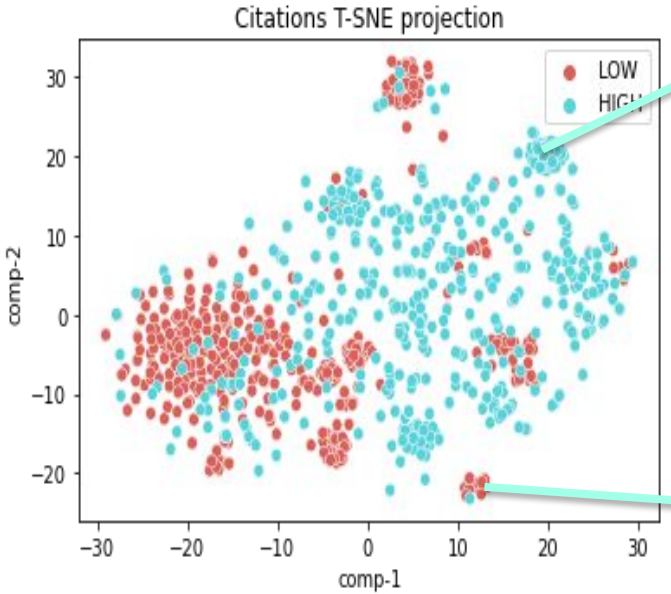
Recall of Hits: Six times greater than base rate

⅔ of all hits are spotted from application data!



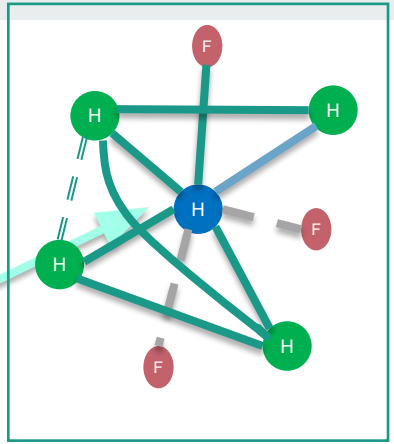
- Our model is a 36% better in producing hits than numerical model

Hidden Structure

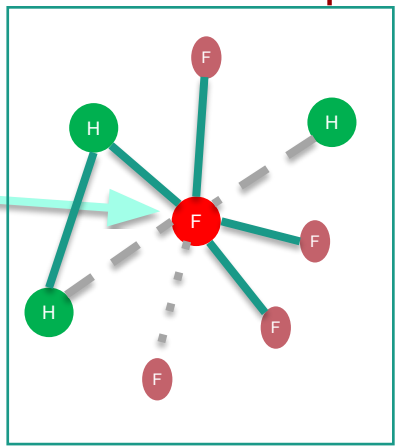


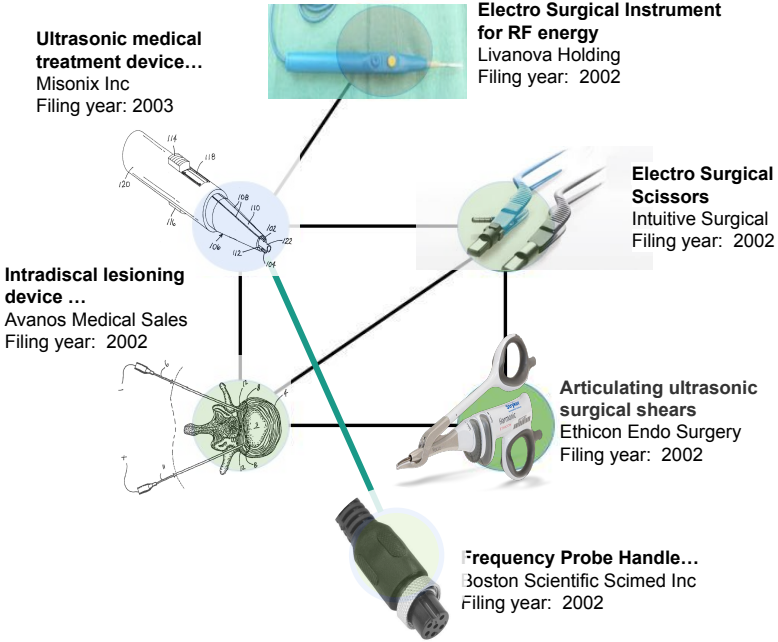
Project in 2D 1K randomly selected Hits and Flops

Network: Hit



Network: Flop

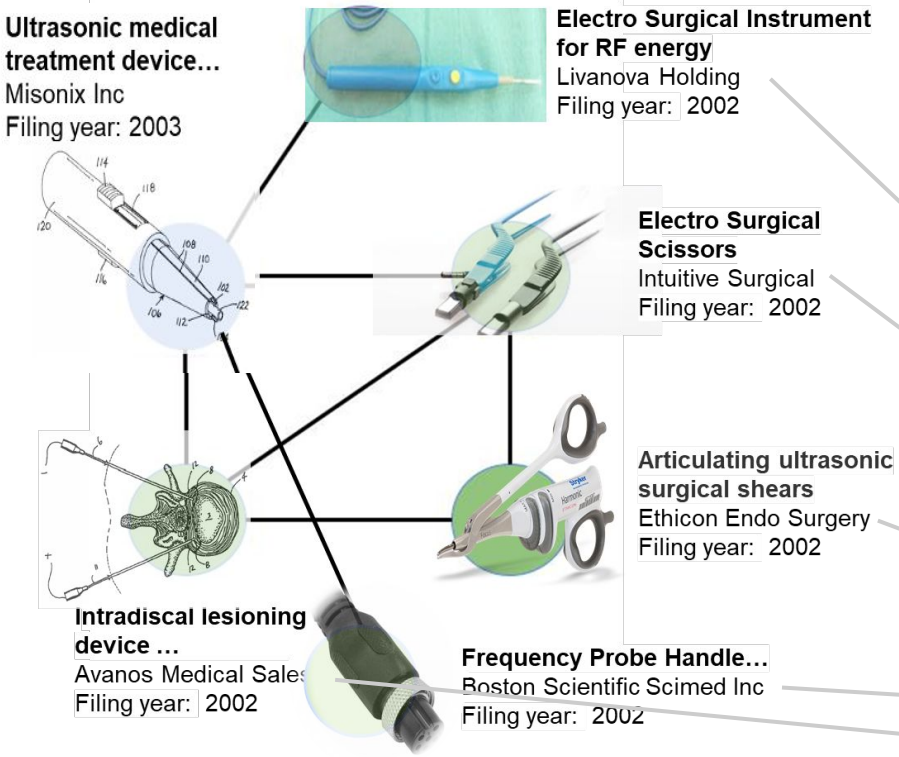




Hit Technologies are part of Embedded Technology Networks

AI identifies Dominant Future Technologies <u>not</u> by their distinctive capabilities but by their codependency with other inventions

Example Embedded Invention Network



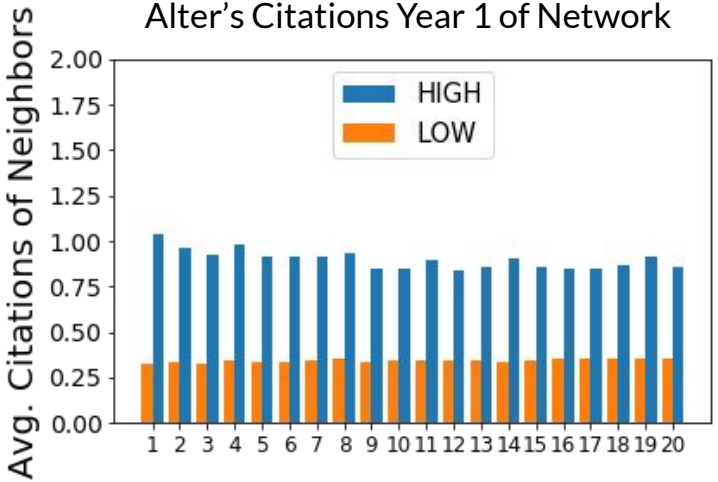
<u>Future</u> hit Inventions relative to future flop invention have *Embedded Tech Networks that have:*

- 1.5x more *entity/company diversity*
- 3x as many alters of a similar *vintage*
- *10x more clustering based on semantic similarity*

Hit Patent	Codependent Patents in Embedded Tech Network	
Misonix Corp (2003)	*Livanova holding*	2002
	Intuitive Surgical.	2002
	Ethicon Endo Surgery	2002
	Boston Scientific Scimed Inc	2002
	Avanos Medical Sales	2002

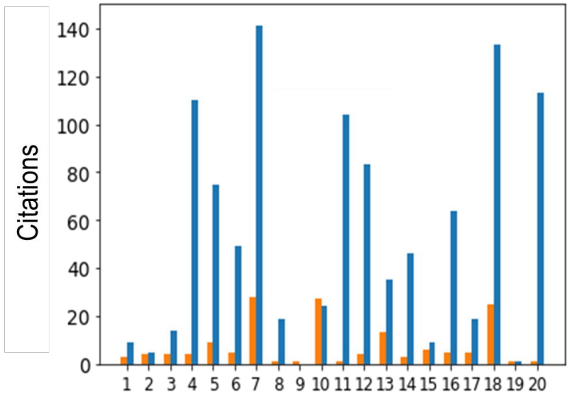
Hit clusters grow together

Alter's Citations Year 1 of Network



Focal Patent's 20 Most Similar Alters

Alter's Citations Year 8 Network



Focal Patent's 20 Most Similar Alters

The neighborhood of a hit patent is more cited than a flop's neighborhood.

Key Takeaways

- Accurate AI model to find hit inventions

- Hidden clusters of impactful technology
    - Clusters are **diverse entities** with **similarly recent birth years**
    - Hit clusters grow together and support future innovation