



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

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

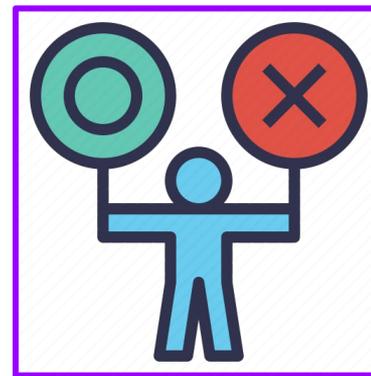
Can we make a model that mimics human decisions on what get patented *before* we predict what becomes hit patents of the future?



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

United States
 (12) Patent Application Publication (40) Pub. No.: US 2018/0373986 A1
 Rainwater (41) Pub. Date: Dec. 27, 2018

(54) MACHINE LEARNING USING DYNAMIC MULTILAYER PERCEPTIONS

(57) U.S. CL. CPC: G06N 3004 (2013.01); G06N 99005 (2013.01); G06N 3404 (2013.01)

(71) Applicant: QbitLogic, Inc., Atlanta, GA (US)

(72) Inventor: Blake Rainwater, Atlanta, GA (US)

(73) Assignee: QbitLogic, Inc., Atlanta, GA (US)

(21) Appl. No.: 15982405

(22) Filed: May 17, 2018

Related U.S. Application Data

(40) Provisional application No. 62524-932, filed on Jan. 26, 2017.

Publication Classification

(51) Int. Cl. G06N 300 (2006.01)
 G06N 344 (2006.01)
 G06N 9900 (2006.01)

Title

Technological Classifications

Abstract

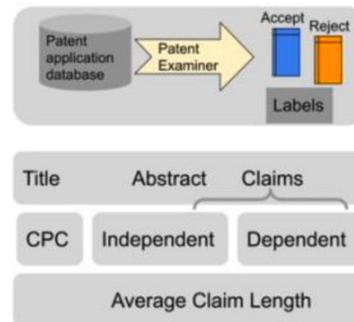
Claim number, type, length, and text

What is claimed is:
 1. A system comprising:
 one or more processors; and
 one or more computer readable media storing instructions, that when executed by the one or more processors, cause the one or more processors to:
 transform first source code into a first abstract syntax tree;
 generate a first dynamic multilayer perceptron network graph based at least in part on the first abstract syntax tree;
 apply a first weight set and the first abstract syntax tree to the first dynamic multilayer perceptron network graph to determine a first calculated output;
 generate a second weight set by adjusting values of the first weight set based at least in part on a comparison of the first calculated output and a first expected output corresponding to the first source code;
 transform second source code into a second abstract syntax tree;
 generate a second dynamic multilayer perceptron network graph based at least in part on the second abstract syntax tree;
 apply the second weight set and the second abstract syntax tree to the second dynamic multilayer perceptron network graph to determine a second calculated output; and
 select the second weight set as a trained weight set based at least in part on a comparison of the second calculated output and a second expected output corresponding to the second source code.

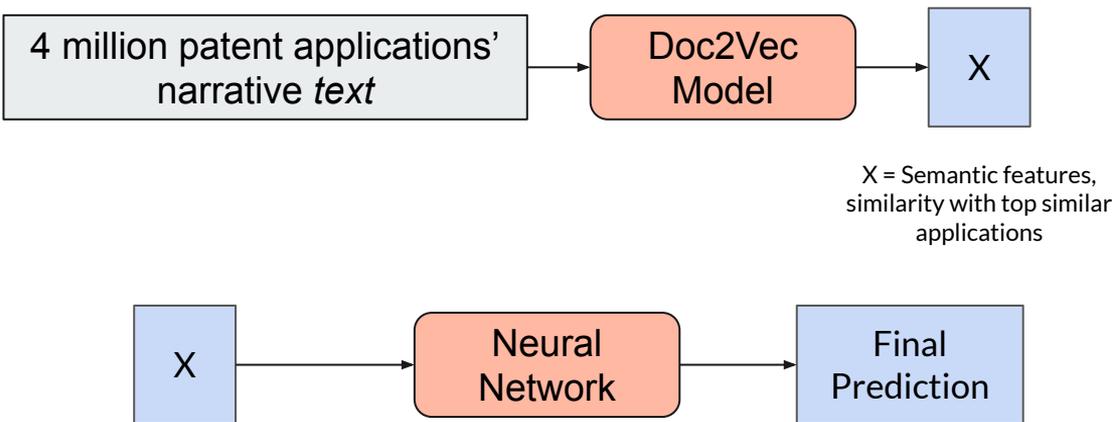
Feature Extraction

- 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

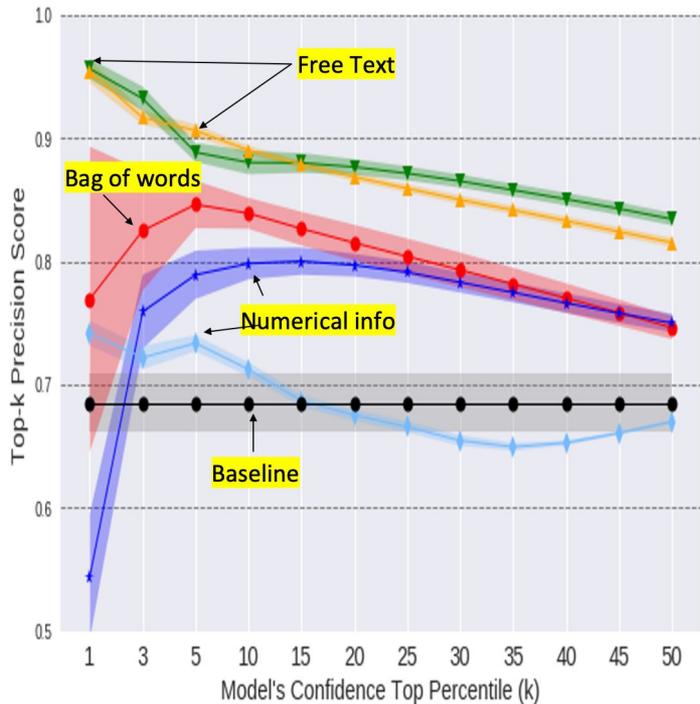


Pipeline



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

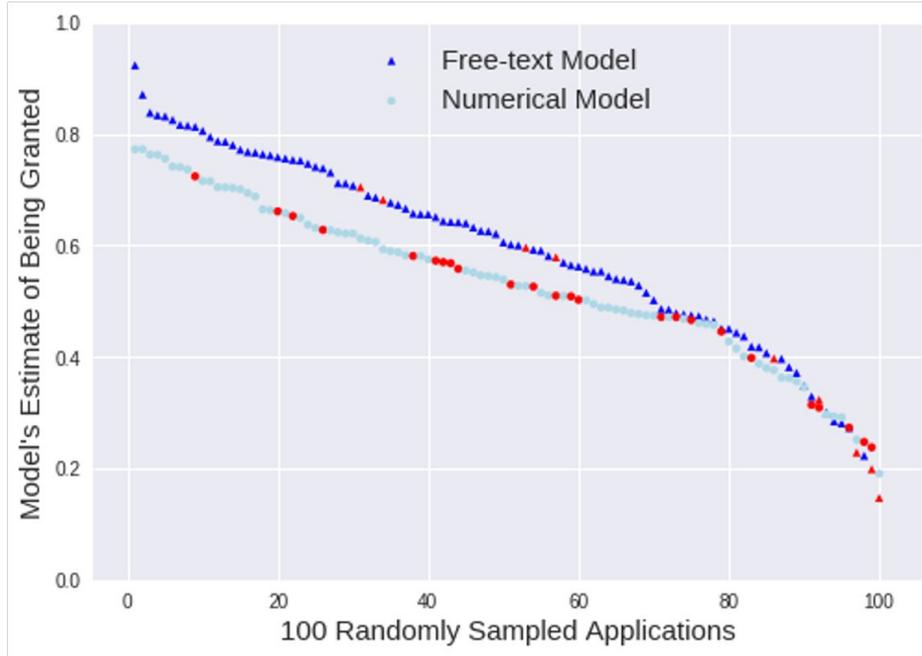


Top K Performance
4 Comparative Models

(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

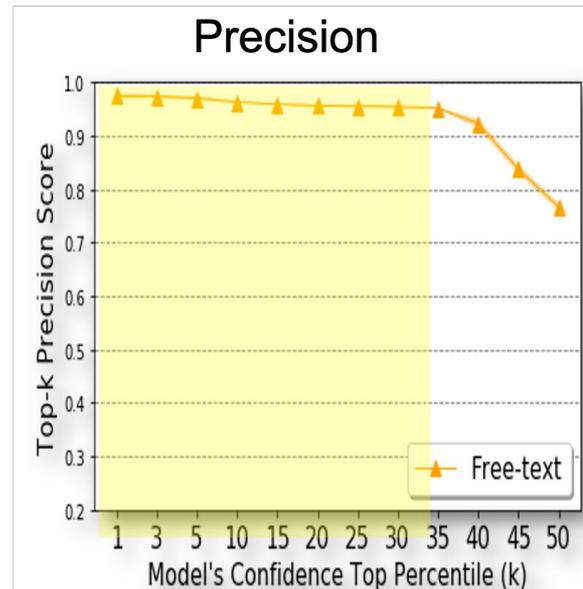
Predict Hit vs Flop Invention

Retrain NN Model

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

Training data: 2001 - 2009

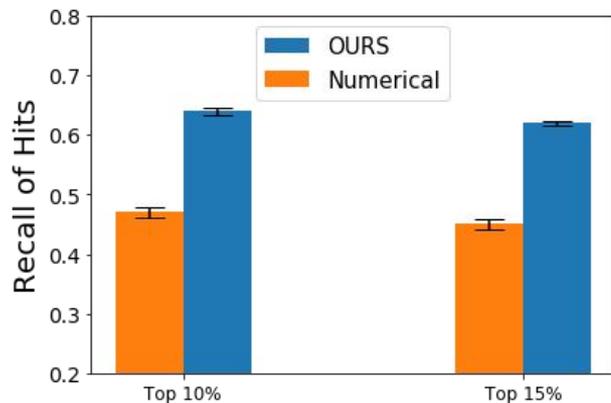
Testing data: 2010 - 2011



Recall of Hits: Six times greater than base rate



$\frac{2}{3}$ of all hits are spotted from application data!

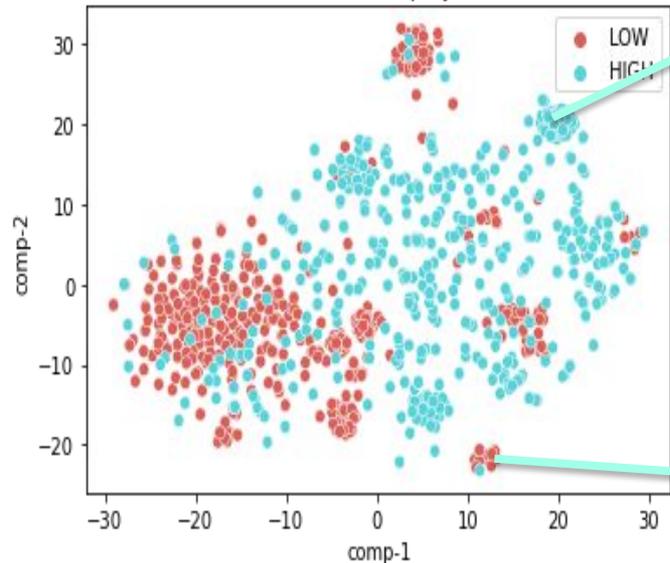


- Our model is a 36% better in producing hits than numerical model

Hidden Structure

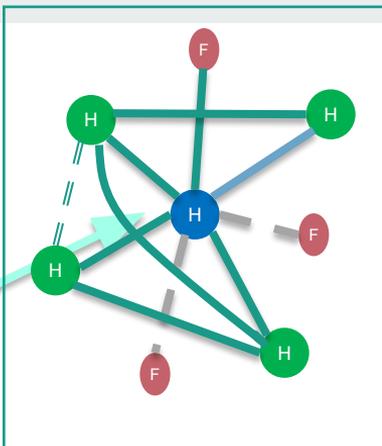


Citations T-SNE projection

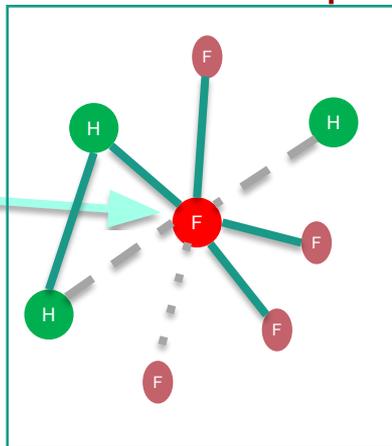


Project in 2D 1K randomly selected Hits and Flops

Network: Hit



Network: Flop

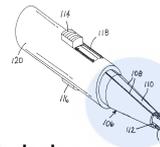


Ultrasonic medical treatment device...
Misonix Inc
Filing year: 2003

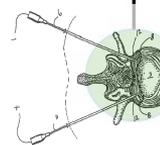


Electro Surgical Instrument for RF energy
Livanova Holding
Filing year: 2002

Intradiscal lesioning device ...
Avanos Medical Sales
Filing year: 2002



Electro Surgical Scissors
Intuitive Surgical
Filing year: 2002



Articulating ultrasonic surgical shears
Ethicon Endo Surgery
Filing year: 2002



Frequency Probe Handle...
Boston Scientific Scimed Inc
Filing year: 2002

Hit Technologies are part of Embedded Technology Networks

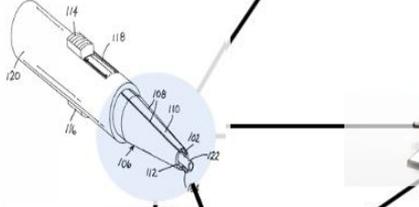
AI identifies Dominant Future Technologies not by their distinctive capabilities but by their codependency with other inventions

Example Embedded Invention Network

Ultrasonic medical treatment device...
Misonix Inc
Filing year: 2003



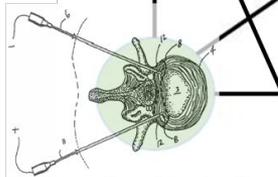
Electro Surgical Instrument for RF energy
Livanova Holding
Filing year: 2002



Electro Surgical Scissors
Intuitive Surgical
Filing year: 2002



Articulating ultrasonic surgical shears
Ethicon Endo Surgery
Filing year: 2002



Intradiscal lesioning device ...
Avanos Medical Sales
Filing year: 2002



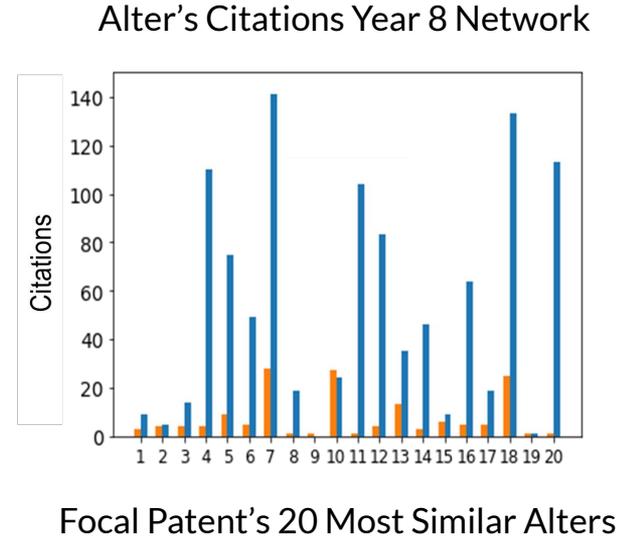
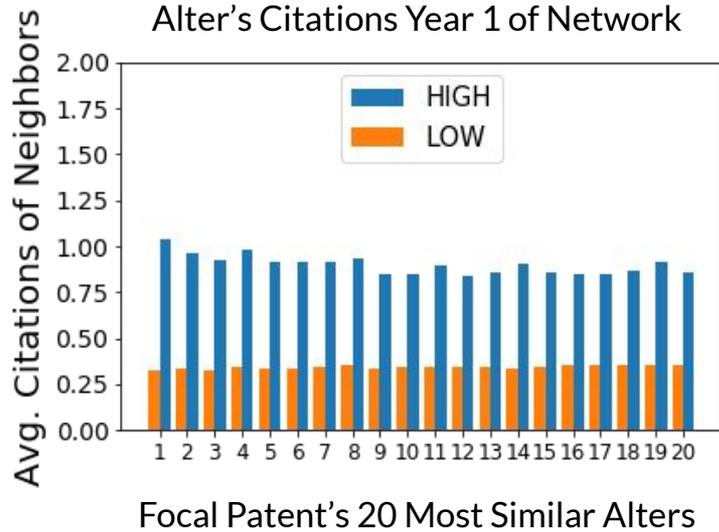
Frequency Probe Handle...
Boston Scientific Scimed Inc
Filing year: 2002

Future 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	
<i>Misonix Corp (2003)</i>	<i>Livanova holding</i>	2002
	<i>Intuitive Surgical.</i>	2002
	<i>Ethicon Endo Surgery</i>	2002
	<i>Boston Scientific Scimed Inc</i>	2002
	<i>Avanos Medical Sales</i>	2002

Hit clusters grow together



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

