

Identifying Patent Monetization Entities



Mihai Surdeanu

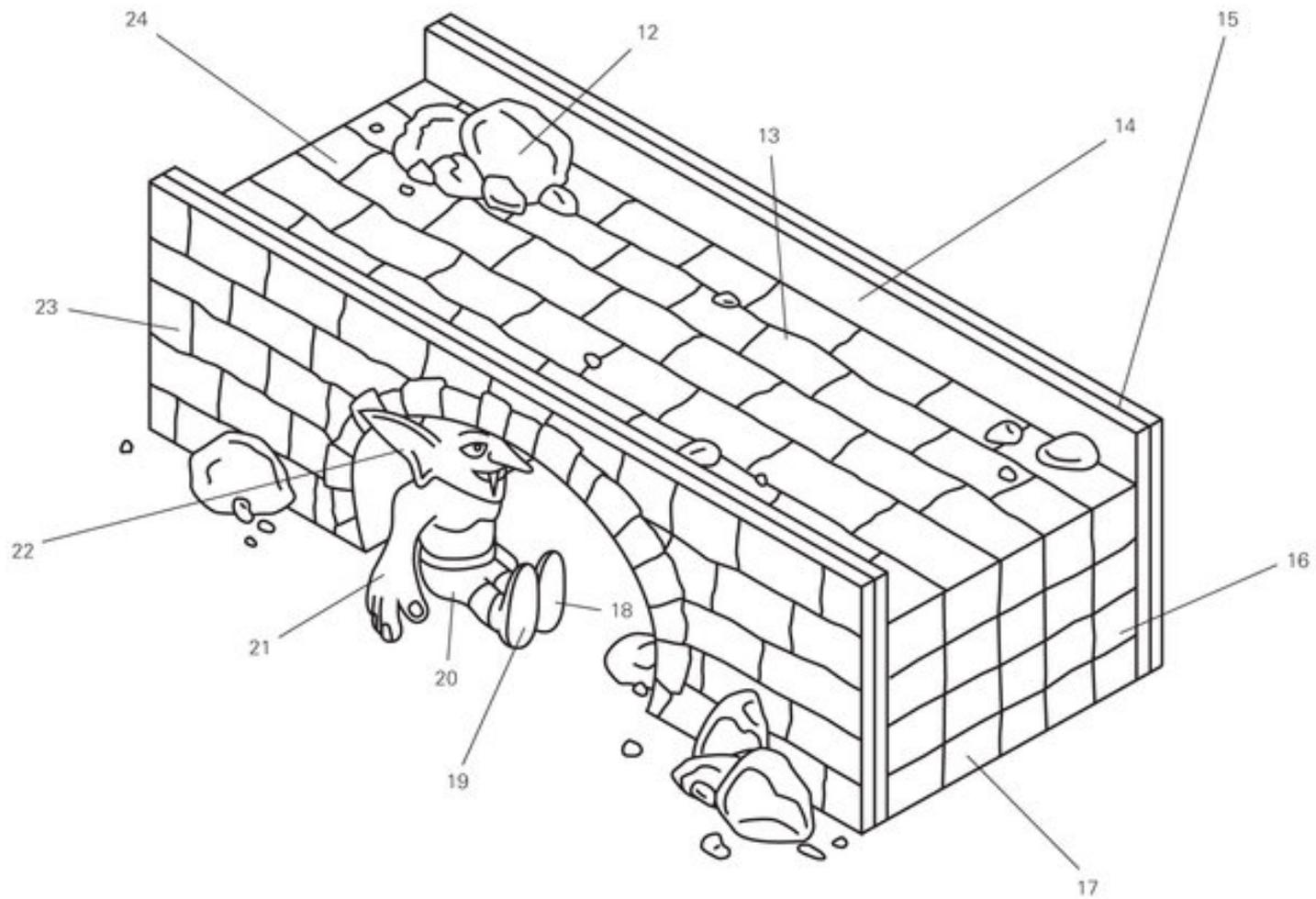
msurdeanu@email.arizona.edu

mihai@lexmachina.com

Sara Jeruss

sjeruss@lexmachina.com

June 13th, 2013



Source: The New York Times, <http://nyti.ms/11qsmVI>

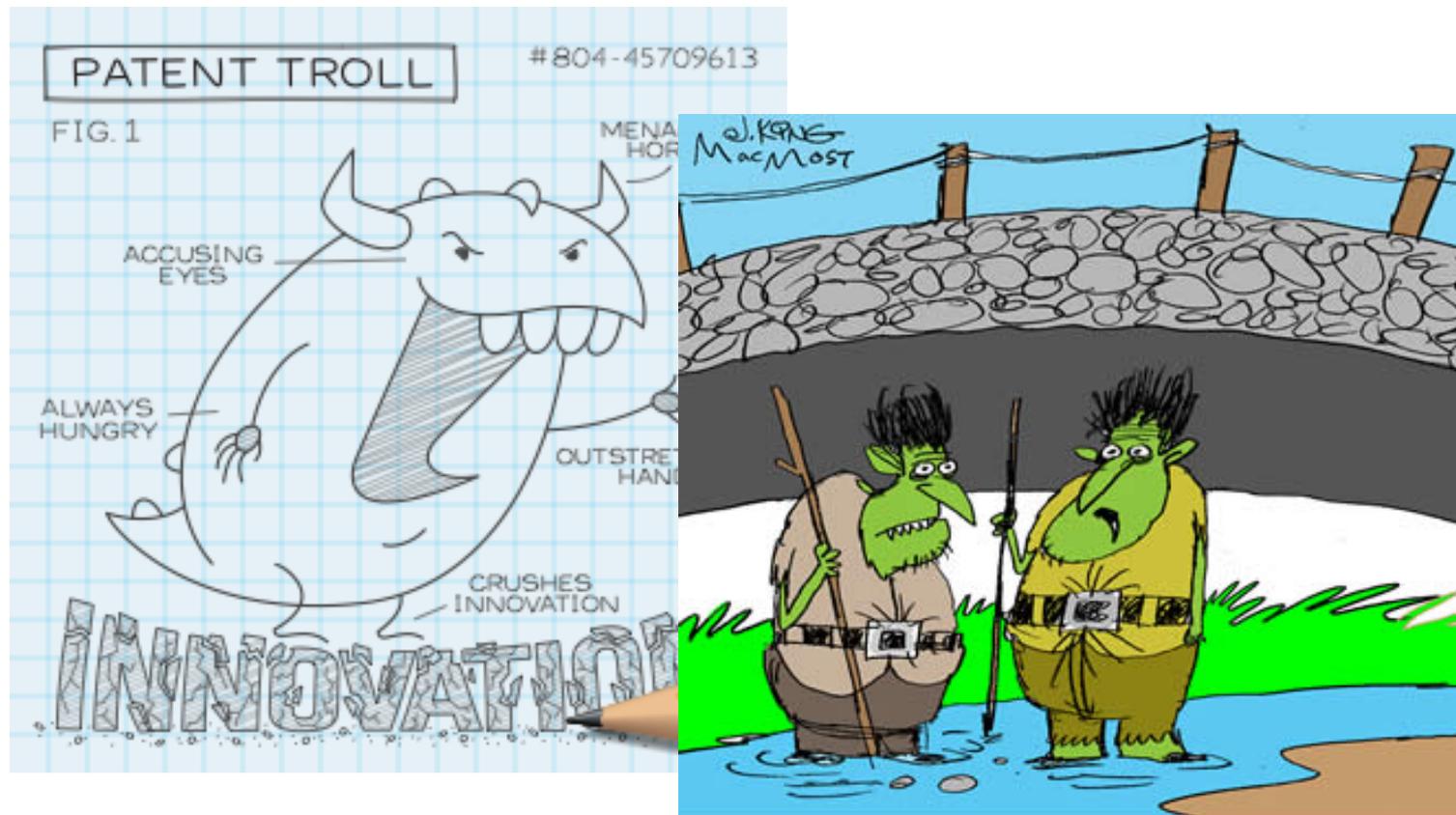
Definition

Patent Monetization Entities (PME) = companies that hold patents, license patents, and file patent lawsuits, but do not sell related products or services.

Other names:

- Patent assertion entities
- Non-practicing entities
- Patent trolls

Rhetoric



“My brother became a patent troll and he now lives under one of those billion-dollar fancy suspension bridges.”

Rhetoric

“Make Patent Trolls Pay in Court”

The New York Times, June 4th, 2013

PMEs just want “to hijack somebody else’s idea and see if they can extort some money.”

President Obama, February 14th, 2013

The House Judiciary Committee
Subcommittee on Courts, Intellectual
Property and the Internet held hearings on
litigation abuse by *“patent trolls”*

March 2013

Rhetoric + Action

In 2011, Congress passed the 2011 Patent Reform Act, aka the “America Invents Act”:

- Study the “*consequences of patent infringement lawsuits brought by non-practicing entities.*”
- Made it harder to file a lawsuit against multiple defendants.

In May 2013, Gov. Shumlin, Democrat of Vermont signed legislation that empowers the attorney general to sue patent holders who assert infringement claims against Vermont businesses or residents in bad faith.

The Problem

We still do not know much about PMEs...

Some fundamental questions:

- How many lawsuits are filed by PMEs?
- Has there been an increase in such lawsuits?
- Is PME litigation behavior and outcomes different from those of other entities?

Some Previous Work

Manual 2012 study conducted by Lex Machina for the “America Invents Act”

- Analyzed 100 cases/year between 2007 and 2011.
- PME lawsuits increased from 22% (in 2007) to 40% (in 2011).

Limitations of Previous Work

- There are no studies that analyzed *all* the patent infringement lawsuits
- Why not?



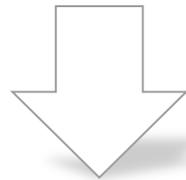
Estimated cost to analyze all patent lawsuits filed between 2000 and 2012: **5+ person years**



Can one analyst maintain consistent accuracy for 5+ years?

Goal of this Work

Automated classification of lawsuit plaintiffs into PMEs or Operating Companies (OC)



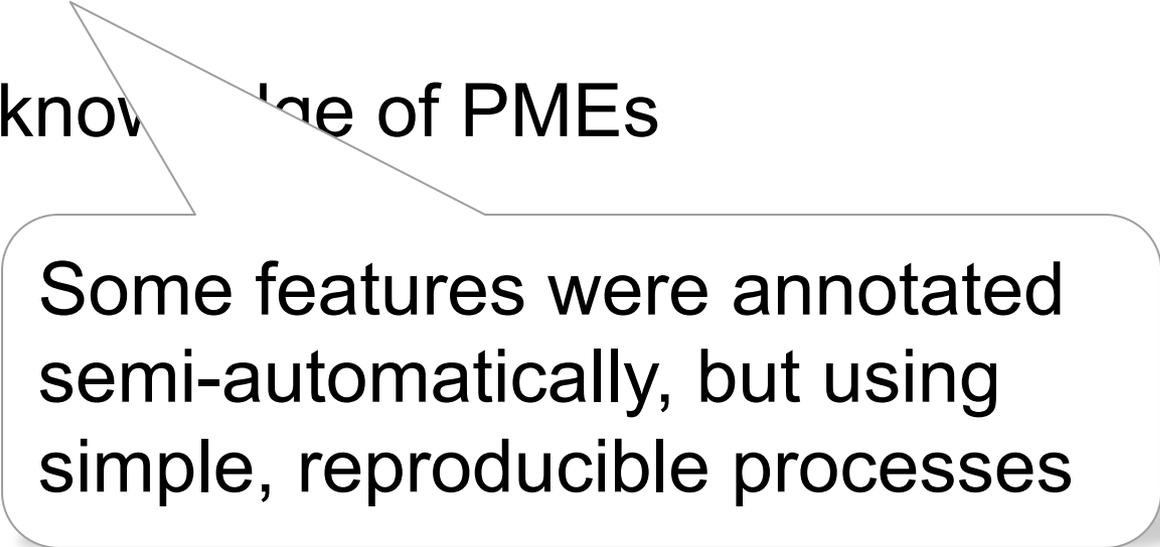
- Gaining a clear picture on the 37,000+ patent lawsuits filed to date
- Allowing policy makers and OCs to assess new lawsuits in real time

Constraints

- Classify plaintiffs in a given lawsuit rather than standalone entities
 - GS Cleantech, 2005: *“development stage company,” which “commercializ[ed] oil extraction technologies”*
 - GS Cleantech, 2010: *“a streamlined, post-market acceptance, technology licensing company...”*
- Use only litigation data (aggregated by Lex Machina) and publicly available information

Plaintiff Features

- From litigation data
- From raw web text
- From litigation/web text, using NLP
- Non-textual
- Pre-existing knowledge of PMEs



Some features were annotated semi-automatically, but using simple, reproducible processes

Plaintiff Features: From Litigation Data

- The current lawsuit has 2+ or 3+ plaintiffs
- The entity in question has been previously sued in a patent case
- The entity has filed 10+, 20+, or 30+ concurrent cases with this lawsuit
- The entity has filed 10+, 20+, or 30+ lawsuits in the same month in the past

Plaintiff Features: From Raw Web Text

Entities often describe themselves or are described by others in web documents.

“Catch Curve, Inc. is an intellectual property development and licensing company focused on communications and messaging technologies based in Atlanta, Georgia.”

“LunarEYE, has developed and patented hardware which, combined with the black box data recorders designed by Salt Lake City-based Independent Witness Inc., allows operators of vehicle fleets – such as BP – to track the vehicles and respond to various situations.”

Algorithm

1. If entity has a website, use its content. Otherwise, fetch the top hits from a search engine.
2. Extract the sentences containing the entity name.
3. Construct bag-of-word features from these sentences, e.g., `containsWord:licensing`.
 1. Feature weight indicates number of occurrences.
4. If no sentences found, add a binary feature to indicate no web presence.

Plaintiff Features: Text, Using NLP

- Entities often describe themselves in litigation documents.
 - If the entity sells a product/service, it is likely to mention it in the “Facts” section of *complaint* documents or *briefings on motions to transfer*.
- But these documents are very verbose, with many repetitions of the entity’s name.
 - Must use only the relevant snippets! Use NLP.

Algorithm

1. Extract complaint and motion to transfer documents using the classifier of Nallapati and Manning (2008).
2. From these documents we extracted the following features:
 1. sellsProduct: if entity name appears in the same sentence with keywords such as: “*development*”, “*manufacture*” (10 overall)
 2. selfDescriptionAsPME: if name appears in the same sentence with phrases such as: “*licensing*”, “*licensees*”, “*does not sell*” (8 overall)
 3. selfDescriptionAsOC: if sellsProduct or name appears in the same sentence with keywords such as: “*provides*”, “*service*” (4 overall)
3. The same algorithm is also applied to all the web sentences previously extracted.

Plaintiff Features: Non-textual

- Does the plaintiff have the same address as its counsel?
- State of incorporation
- Was the entity incorporated within six months of lawsuit filing date?
- Were the asserted patents assigned to this entity within 6 months of lawsuit filing date?
- Does the entity have a website?

Plaintiff Features: Pre-existing Knowledge of PME

- Lex Machina already has a DB of known PMEs and law firms known to represent PMEs
- Were the patents assigned to this entity by a known PME?
- Is the entity's counsel known to represent PMEs?

Model

- **Logistic regression**
- L2 regularization
- L-BFGS optimization
- We did not tune σ , the regularization parameter

Data

- 400 lawsuit plaintiffs, randomly selected from lawsuits filed in 2007
 - We eliminated 30 lawsuits, where the plaintiffs could not be reliably classified by annotators.
 - 370 lawsuits with 353 unique plaintiffs
- All annotations were created by law students and were reviewed by one of the authors.
- 73% of lawsuits were initiated by OCs.
- Five-fold cross validation experiments

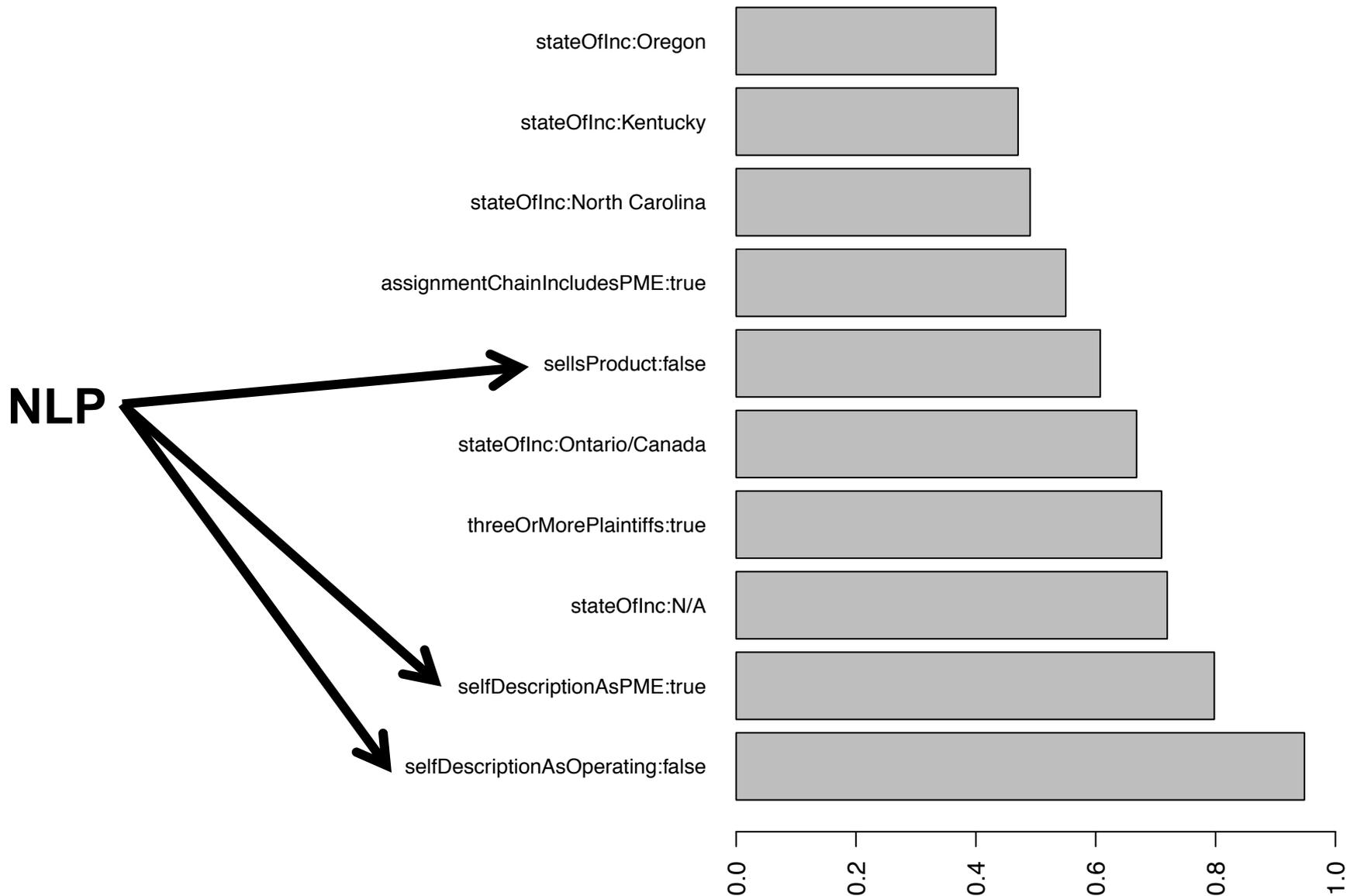
Evaluation Metrics

- Accuracy = correct predictions / total predictions
- Precision (P) = correct PME predictions / total PME predictions
- Recall (R) = correct PME predictions / total PME lawsuits
- $F1 = 2PR / (P + R)$

Results

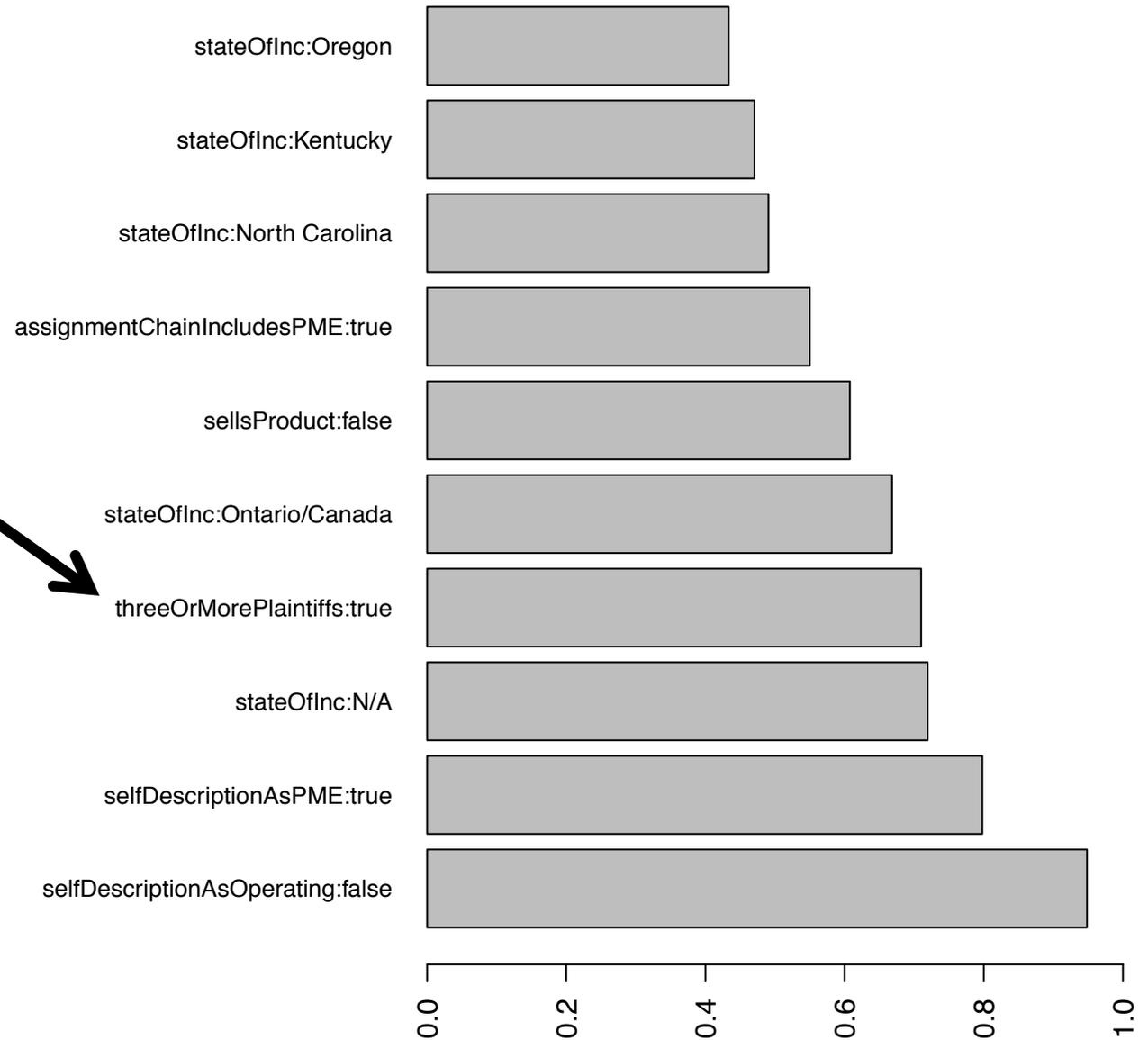
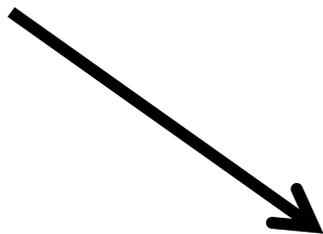
	Accuracy	Precision	Recall	F1
Baseline	72.8	–	–	–
Complete	92.2	87.5	83.2	85.3
– NLP	82.7	70.8	62.4	66.3*
– non-textual	90.8	85.3	80.2	82.7*
– litigation data	91.4	88.8	78.2	83.2*
– raw web text	92.2	89.1	81.2	85.0*
– knowledge of PMEs	92.2	87.6	83.2	85.3

Top 10 Features for PME

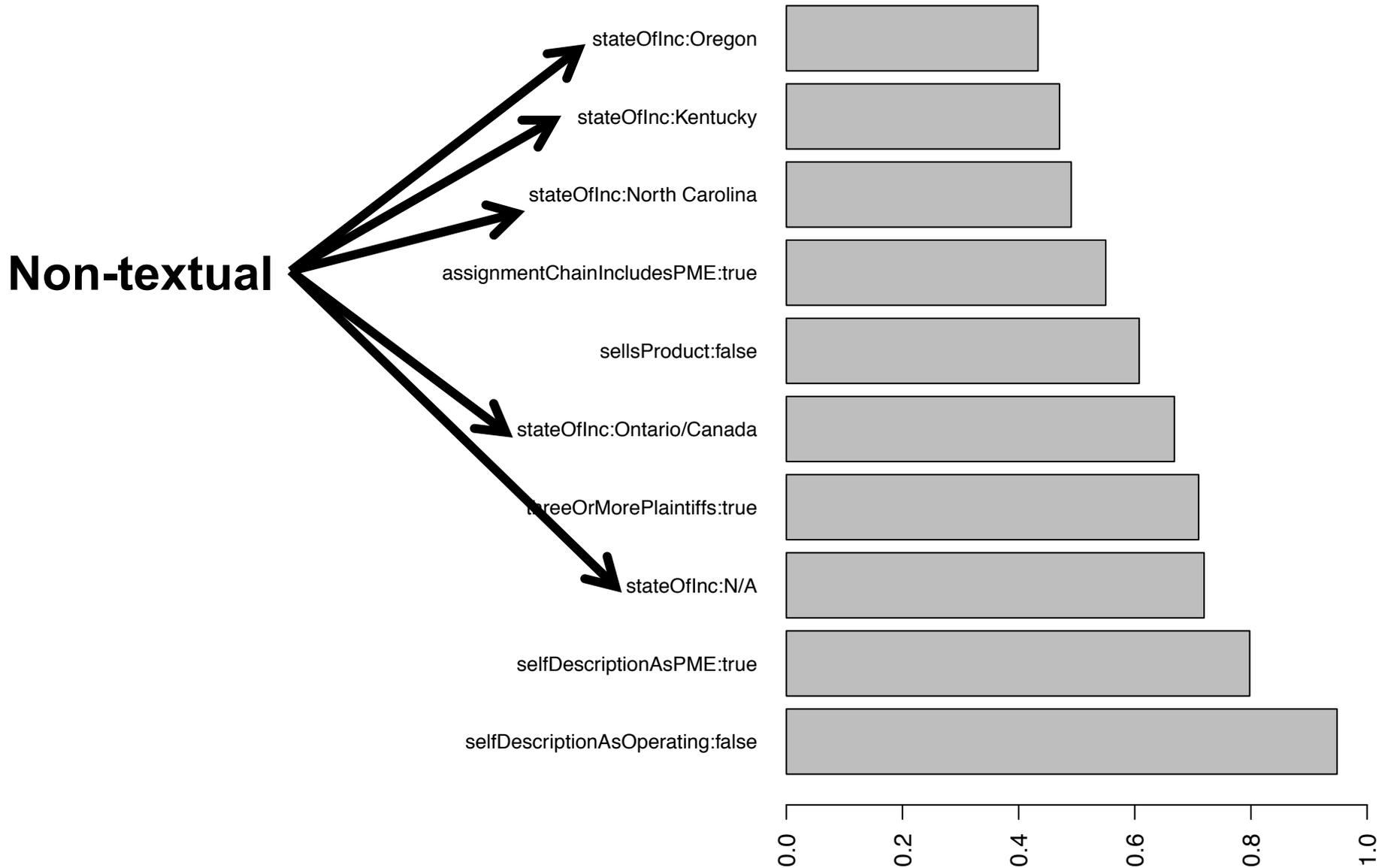


Top 10 Features for PME

Litigation

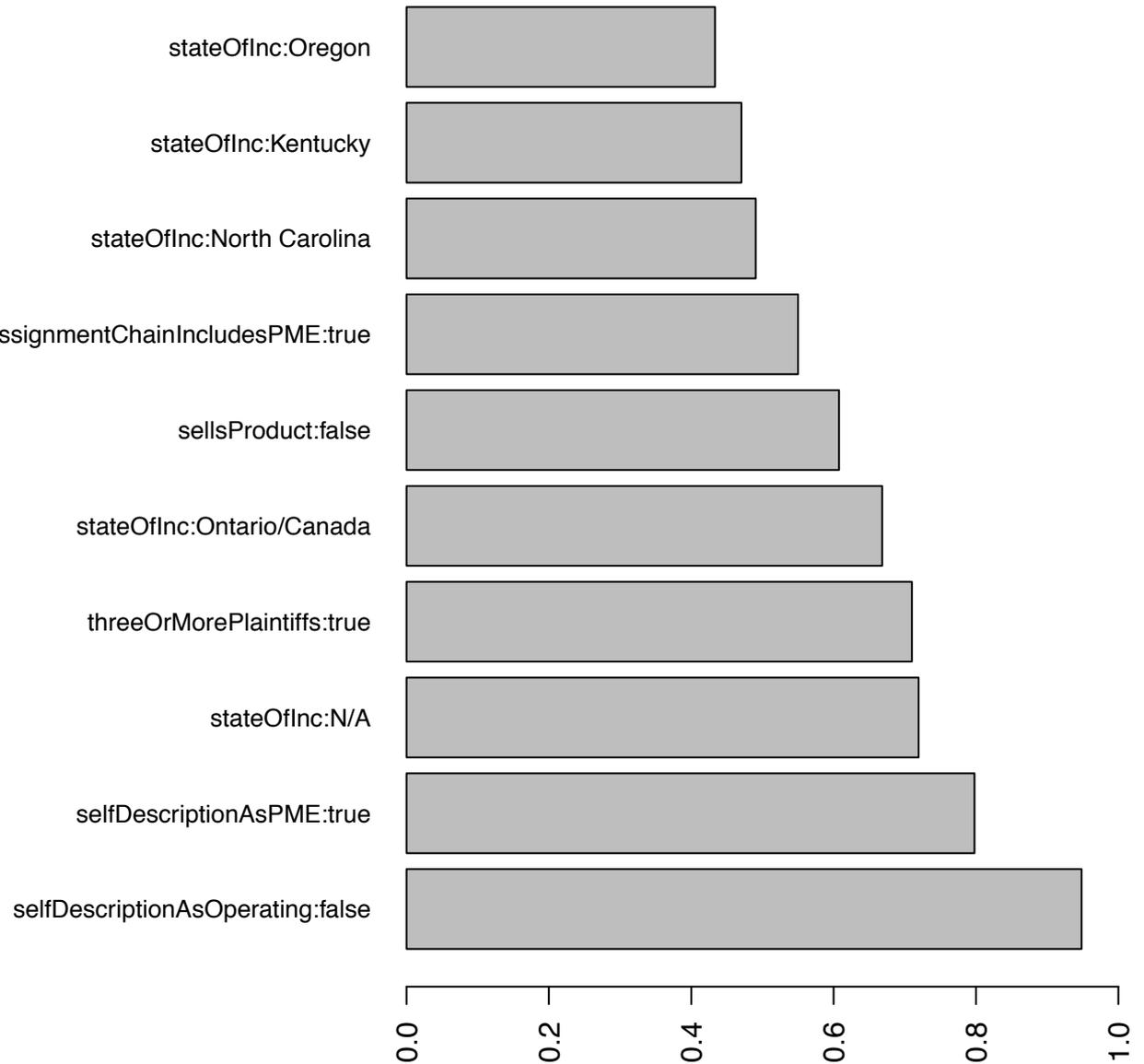
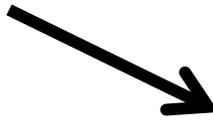


Top 10 Features for PME

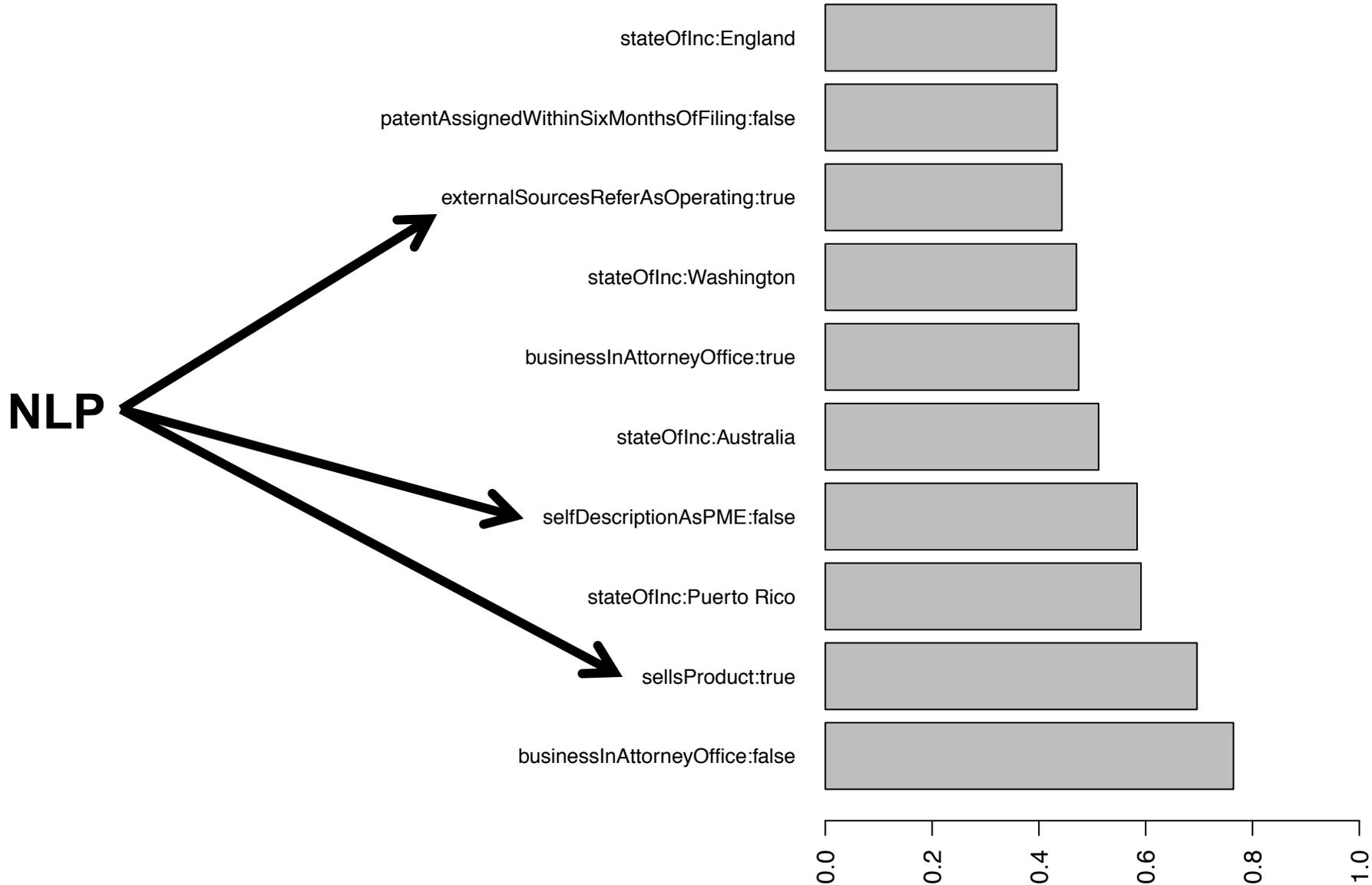


Top 10 Features for PME

**PME
Knowledge**

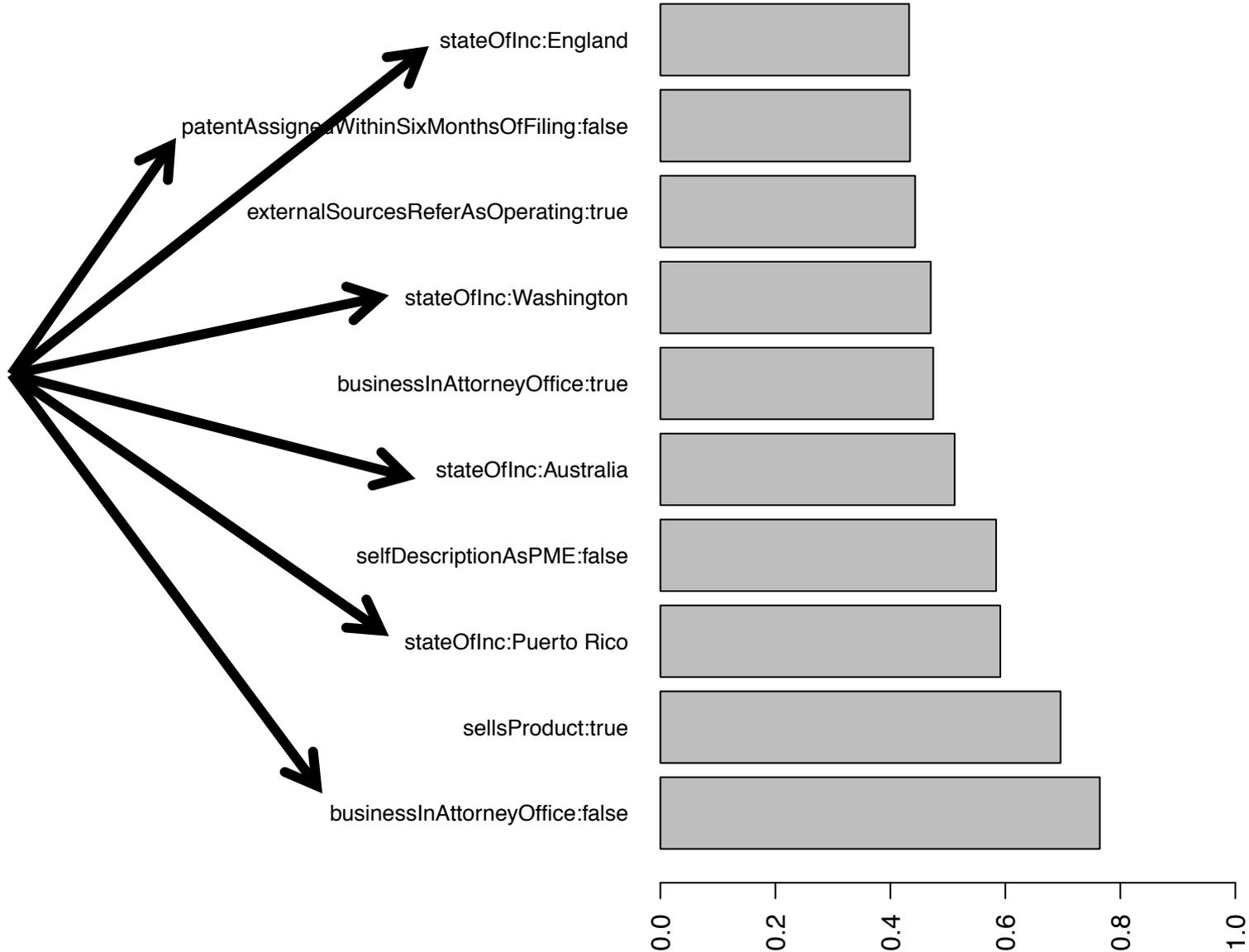


Top 10 Features for OCs



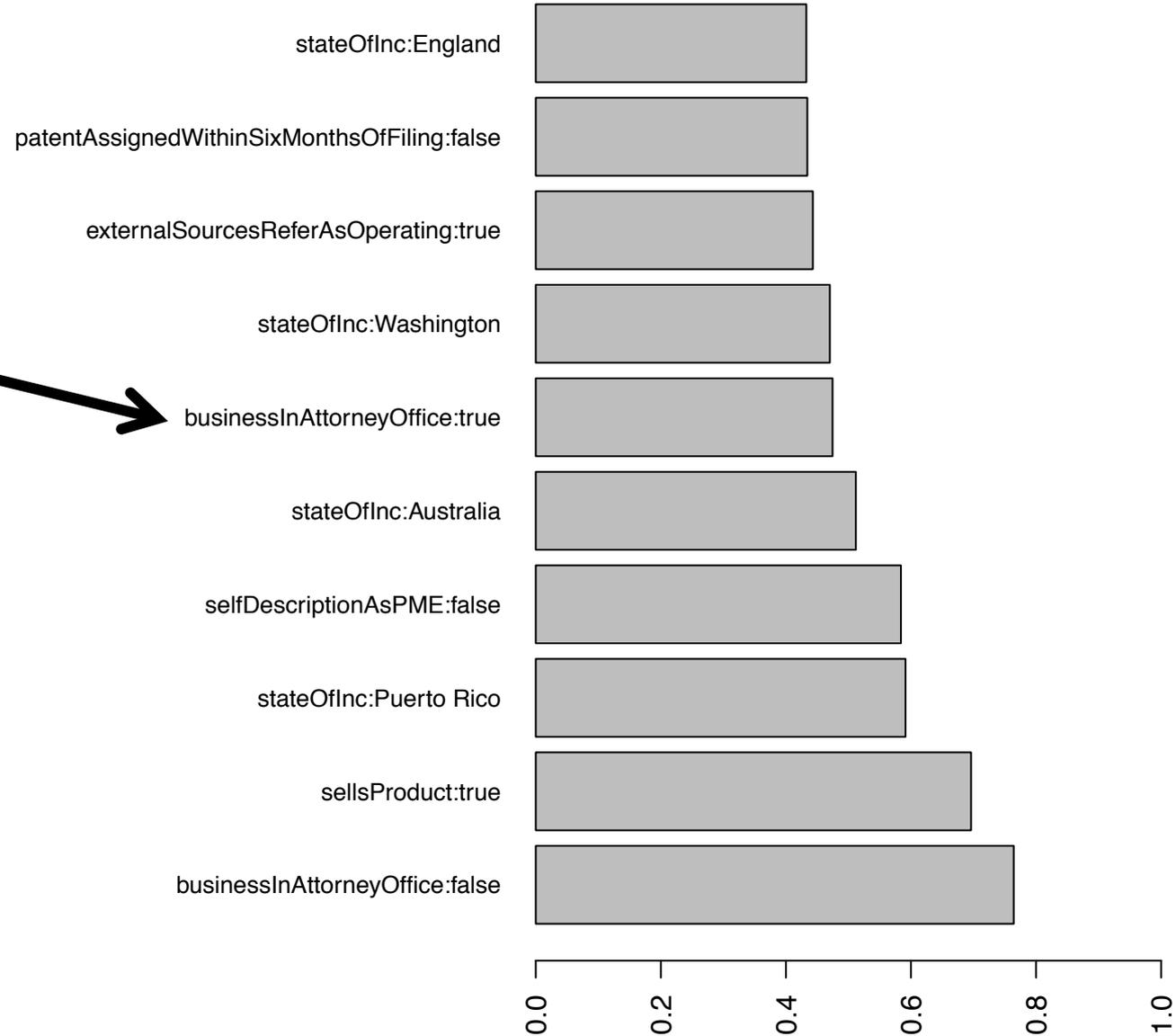
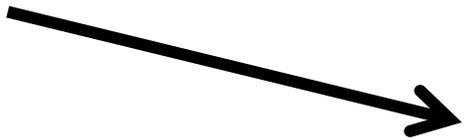
Top 10 Features for OCs

Non-textual



Top 10 Features for OCs

Overfit!



Error Analysis of False Negatives

- 50% annotation errors
 - Annotators confused Monsanto Company with Monsanto Technology LLC
- 50% complex entities, with properties of both OCs and PMEs
 - Wake Forest Health Sciences
 - University division + focus on monetization
 - Bear Creek Technologies
 - 2005: “IT company specializing in the development of software solutions, automated software products, and technological services...”
 - Opposition to motion to transfer, 2011: all development and sales activity in the past
 - Website stale since 2005

Conclusions

- First empirical model for the identification of patent monetization entities
- Extracts PME's with an F1 score of 85%
- A strong case for the utility of NLP in the legal domain

Acknowledgments

- Robin Feldman, Joshua Walker

} **insights**

- John Beard, Nicholas Billings, Padmini Cheruvu, Byron Huang, Tasha Iyer, Umar Khan, Rachel Kinney, Adrian Kwan, Brian Lee, Jagtej Sodhi, John Tynes, and Yi Wilkinson

} **annotations**

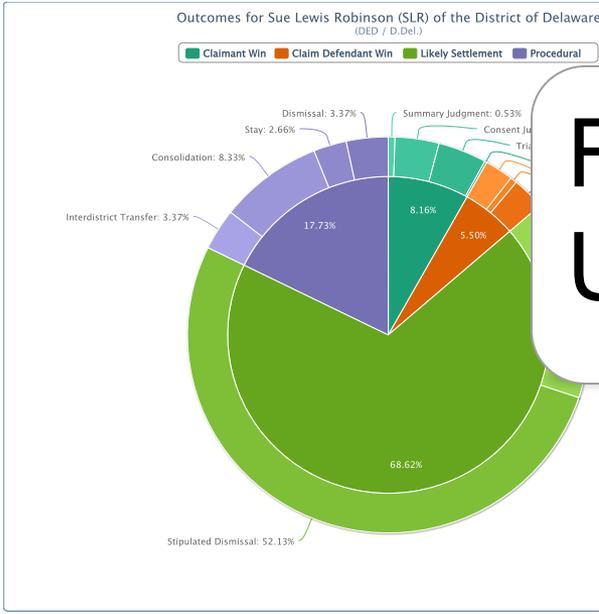
- Lex Machina

} **data**

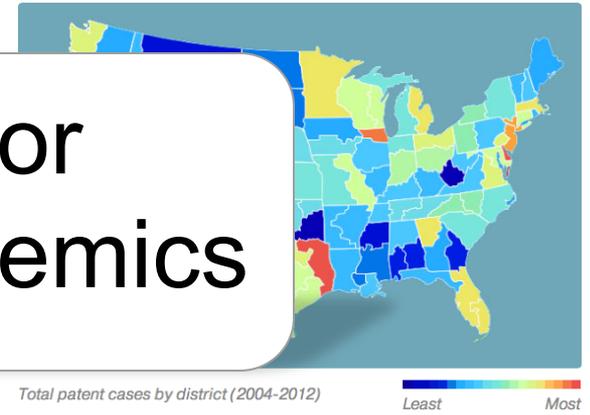
Lex Machina

IP Litigation Data and Analytics

Judge Robinson – Patent Case Outcomes



Total Patent Cases by District



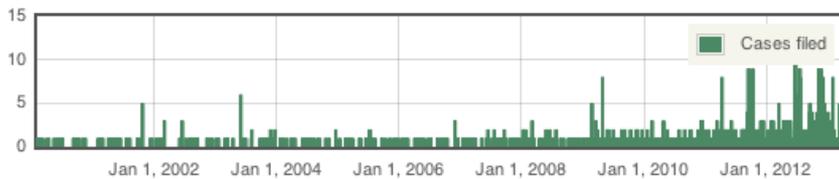
Free accounts for US-based academics

Open Cases Over Time

Samsung All Cases – Time to Termination

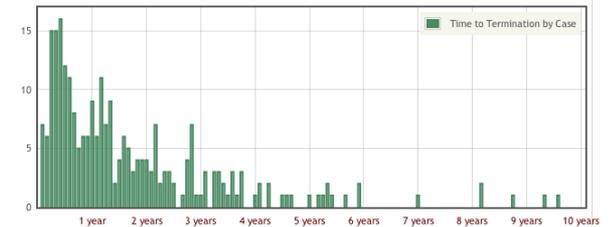
Jack Blumenfeld – Cases Filed

Cases Filed (910 potentially related cases)



Termination Times

Number of Cases	257
Mean	557 days
Range	2 - 3202 days





THANK YOU! QUESTIONS?