

(min) Forskningsstøtte i Fremtiden

Anders Søgaard

<http://anderssoegaard.github.io>



coASTal



UNIVERSITY OF
COPENHAGEN



Indhold

- Lidt om NLP
- Hvordan får jeg penge?
- Hvorfor gaver?
- **Fond-forsker matching**
- Nye krav til forskningssupport
- Målbar impact?



Innovationsfonden

facebook®

CARISBERG FOUNDATION



Joachim Bingel

Postdoc
NLP, Dialogue, Simplification



Ana Valeria Gonzalez

PHD Student
NLP, Dialogue



Mareike Hartmann

PHD Student
NLP, Social Science



Desmond Elliott

Assistant Professor
NLP, Dialogue



Victor Petren Hansen

PHD Student
NLP, Dialogue



Maria Barrett

Postdoc
NLP, Dialogue



Yova Kementchedjheva

PHD Student
NLP, Unsupervised Learning



Mostafa Abdou

PHD Student
NLU, Semantic Parsing



Anders Søgaard

Professor
NLP, Machine Learning



Simon Flachs

PHD Student
NLP, Error correction



Rahul Aralikkatte

PHD Student
NLU, Machine Learning



Daniel Hershovich

Postdoc
NLP



Heather Lent

PHD
NLP



Lasse Borgholt

PHD
NLP, Speech



Marcel Bollmann

Postdoc
NLP, Morphology



Lidt om NLP



- Tokenization
- Lemmatisation
- POS tagging
- WSD
- MWE identification
- NER
- Chunking
- Dependency parsing
- Constituent parsing
- Coreference resolution
- Semantic parsing
- Discourse parsing

- Document classification
- Forecasting, diagnosis support, etc.
- Search and extraction
- Relation extraction, knowledge base population
- Text-to-speech and speech-to-text
- Text simplification
- Summarization
- Question-answering
- Dialogue systems
- Machine translation
- Discourse parsing



- Tokenization
- Lemmatisation
- POS tagging
- WSD
- MWE identification
- NER
- Chunking
- Dependency parsing
- Constituent parsing
- Coreference resolution
- Semantic parsing
- Discourse parsing

- Document classification
- Forecasting, diagnosis support, etc.
- Search and extraction
- Relation extraction, knowledge base population
- Text-to-speech and speech-to-text
- Text simplification
- Summarization
- Question-answering
- Dialogue systems
- Machine translation
- Discourse parsing

Q: Hvordan gør man?

Q: Hvordan gør man?

A: Ligesom i folkeskolen...

x

y

4

5

6

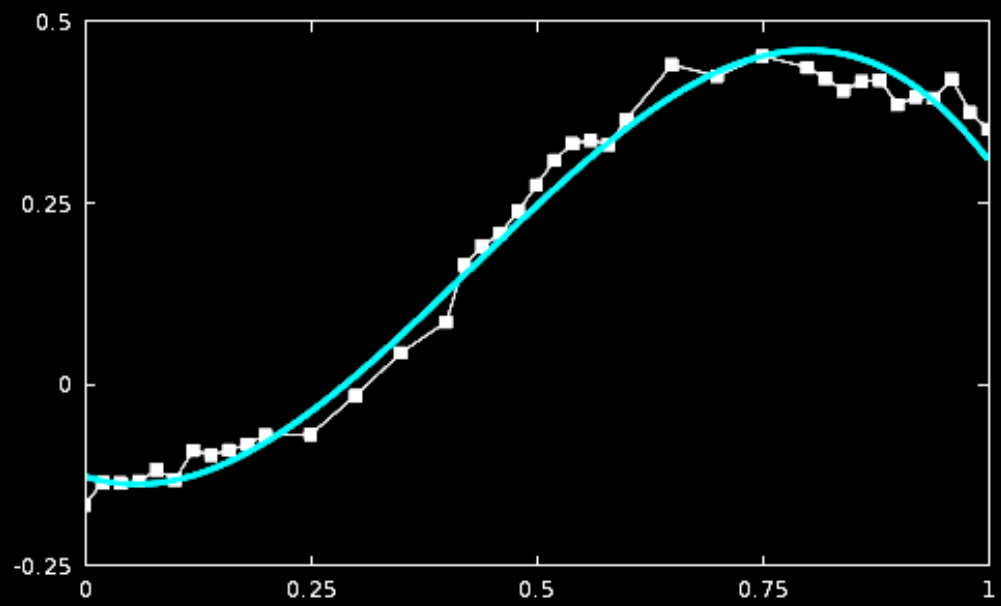
7

11

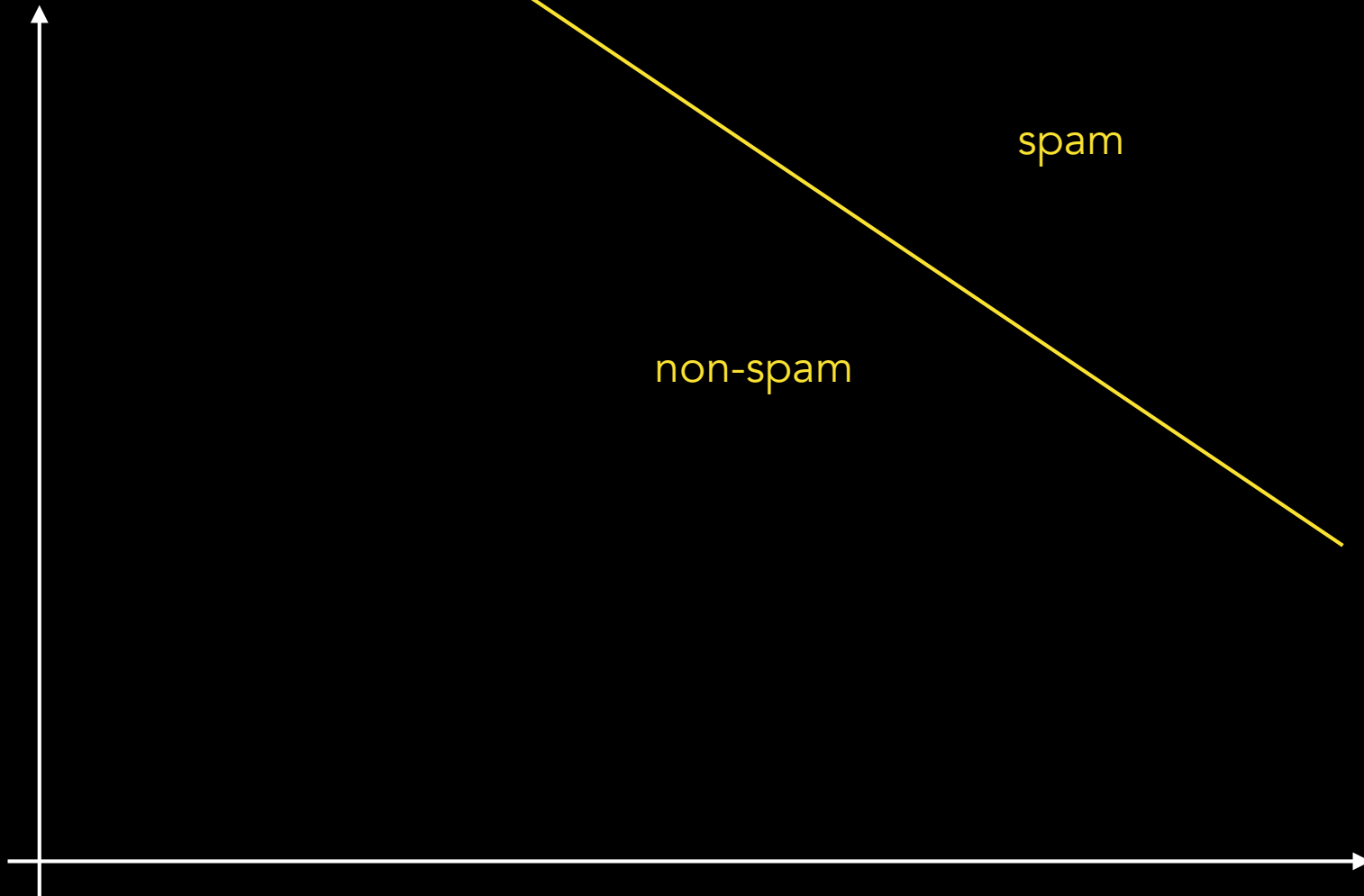
12

13

14



lottery



spam

non-spam

won

"It's busy times. Time flies like arrows."
=> [NOUN, VERB, ADV, NOUN]

"Banana flies like fruits. Time flies like arrows."
=> [NOUN, NOUN, VERB, NOUN]



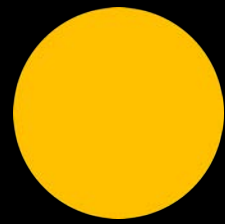
- Tokenization
- Lemmatisation
- POS tagging
- WSD
- MWE identification
- NER
- Chunking
- Dependency parsing
- Constituent parsing
- Coreference resolution
- Semantic parsing
- Discourse parsing

- Document classification
- Forecasting, diagnosis support, etc.
- Search and extraction
- Relation extraction, knowledge base population
- Text-to-speech and speech-to-text
- Text simplification
- Summarization
- Question-answering
- Dialogue systems
- Machine translation
- Discourse parsing

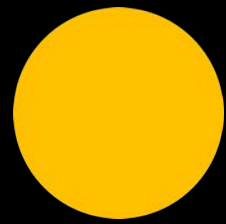
Hvordan får jeg
penge?

	Public	Private Foundation	Private Company
Innovation Fund	Public		
Innovation Fund			
Innovation Fund			
Innovation Fund			
Calsberg		Private Foundation	
Lundbeck			
Trygfonden			
BotXO			Private Company
Telefonica			
Microsoft			
Amazon			
Google			
Facebook			
AI2			

	Public	Private Foundation	Private Company
Innovation Fund	[Blue Block]		
Innovation Fund			
Innovation Fund			
Innovation Fund			
Calsberg		[Green Block]	
Lundbeck			
Trygfonden			
BotXO			[Yellow Circle]
Telefonica			[Purple Bar]
Microsoft			[Yellow Circle]
Amazon			[Yellow Circle]
Google			[Yellow Circle]
Facebook			[Yellow Circle]
AI2			[Yellow Circle]

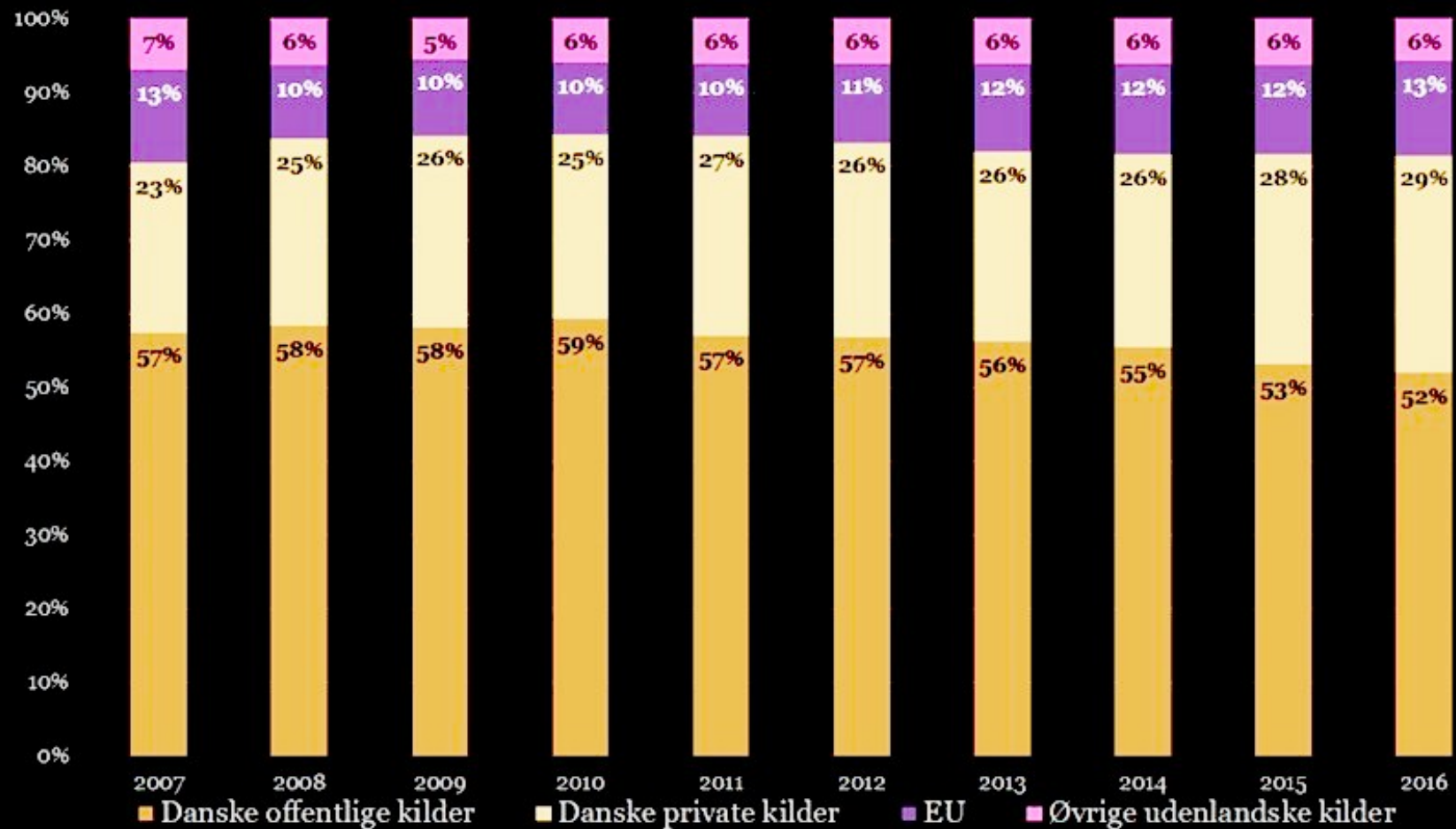
 = No call

	Public	Private Foundation	Private Company
Innovation Fund	■		
Innovation Fund			
Innovation Fund			
Innovation Fund			
Calsberg		■	
Lundbeck			
Trygfonden			
BotXO			■
Telefonica			●
Microsoft			●
Amazon			●
Google			●
Facebook			●
AI2			●



= Unrestricted gift

Fordeelingen af eksterne kilder til finansiering af forskning på universiteter



Hvorfor gaver?

- Jagt på talent
- Synlighed på konferencer
- Branding vha pressedækning af forskningsresultater
- Charity

Fellowships

Google Doctoral Fellowship

Amazon Alexa Fellowship

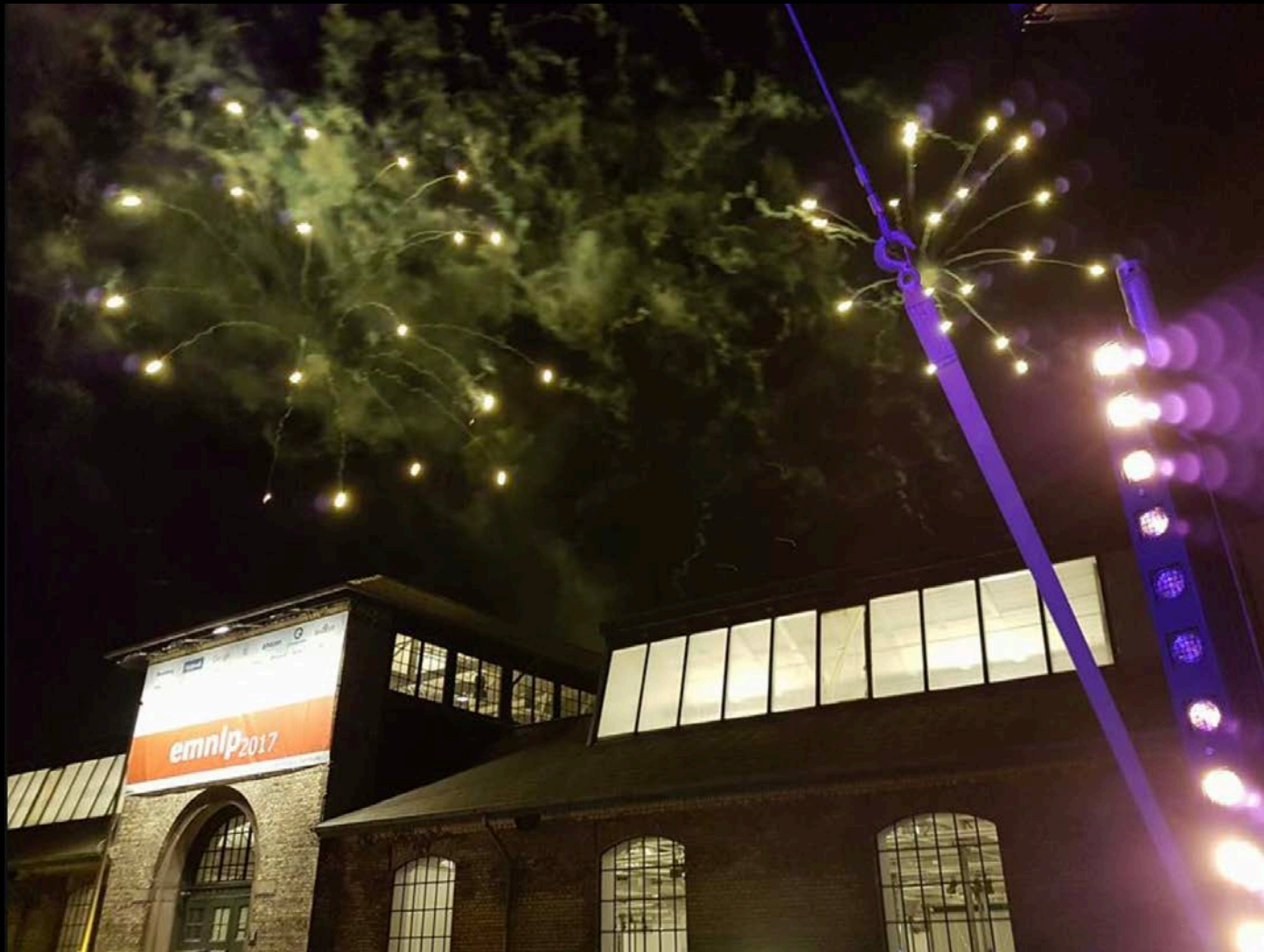
Microsoft PhD Fellowship

Ada Lovelace Fellowship

Bloomberg PhD Fellowships

...

- Jagt på talent
- Synlighed på konferencer
- Branding vha pressedækning af forskningsresultater
- Charity



- Jagt på talent
- Synlighed på konferencer
- Branding vha pressedækning af forskningsresultater
- Charity

MICROSOFT \ TECH \ ARTIFICIAL INTELLIGENCE \

Microsoft reaches 'human parity' with new speech recognition system

By [Sam Byford](#) | [@345triangle](#) | Oct 18, 2016, 9:47pm EDT

- Jagt på talent
- Synlighed på konferencer
- Branding vha pressedækning af forskningsresultater
- Charity

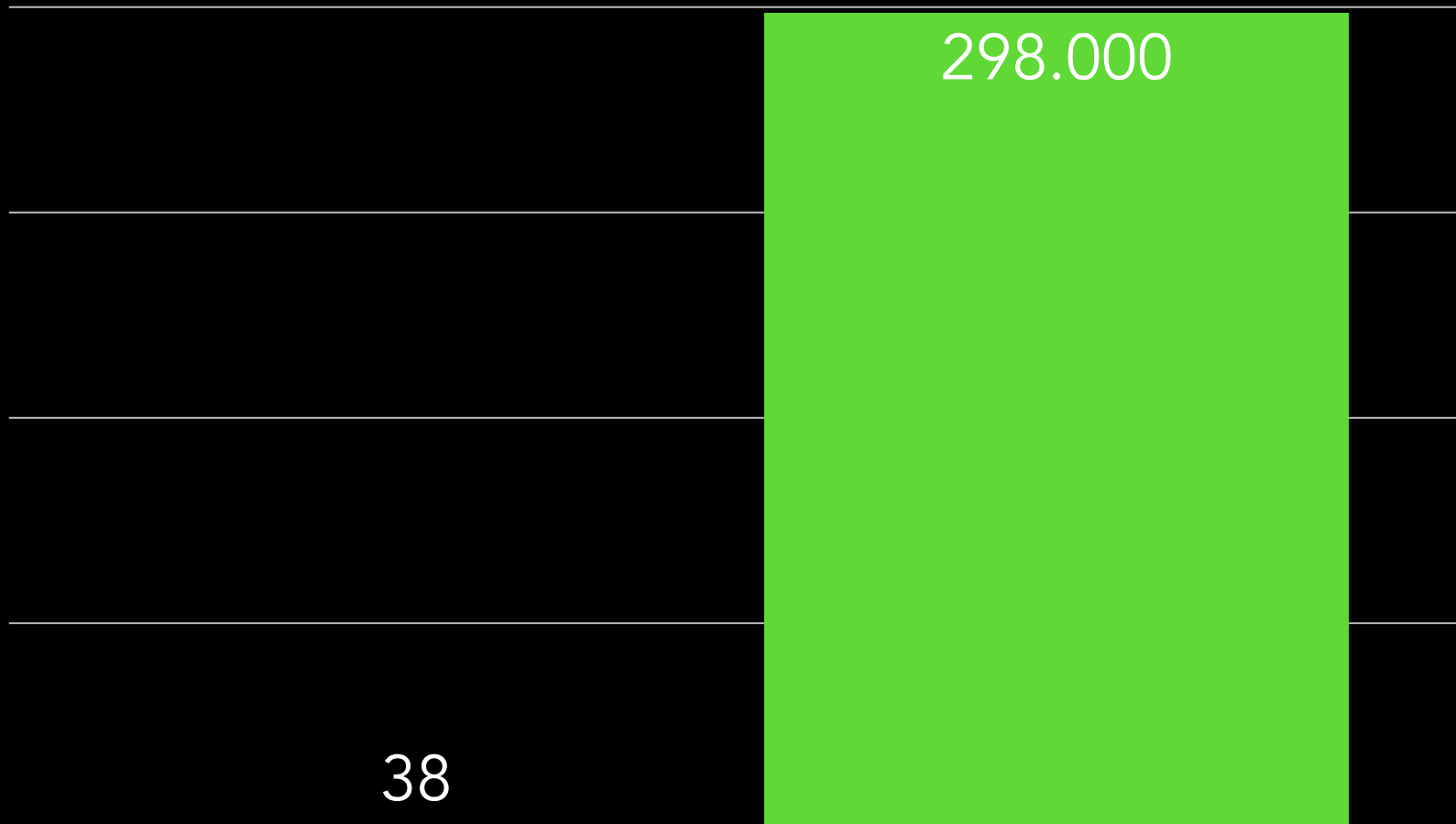
Fond-forsker matching



Fonde



Virksomheder



Forudsigelse af fond-forsker match

Spørgsmål-svar-matching

Tan et al. (2016)

Tekstuel inferens

Schönmackers et
al. (2009)

Tekstbaserede recommender-systemer

Industry

Jobopslag-kandidat-matching

White et al.
(2012)

Forudsigelse af fond-forsker match

Spørgsmål-svar-matching

Tan et al. (2016)

Tekstuel inferens

Schönmackers et al. (2009)

Tekstbaserede recommender-systemer

Industry

Jobopslag-kandidat-matching

White et al. (2012)

Fond-forsker match

?

Forudsigelse af fond-forsker match

Spørgsmål-svar-matching

Tan et al. (2016)

Tekstuel inferens

Schönmackers et
al. (2009)

Tekstbaserede recommender-systemer

Industry

Jobopslag-kandidat-matching

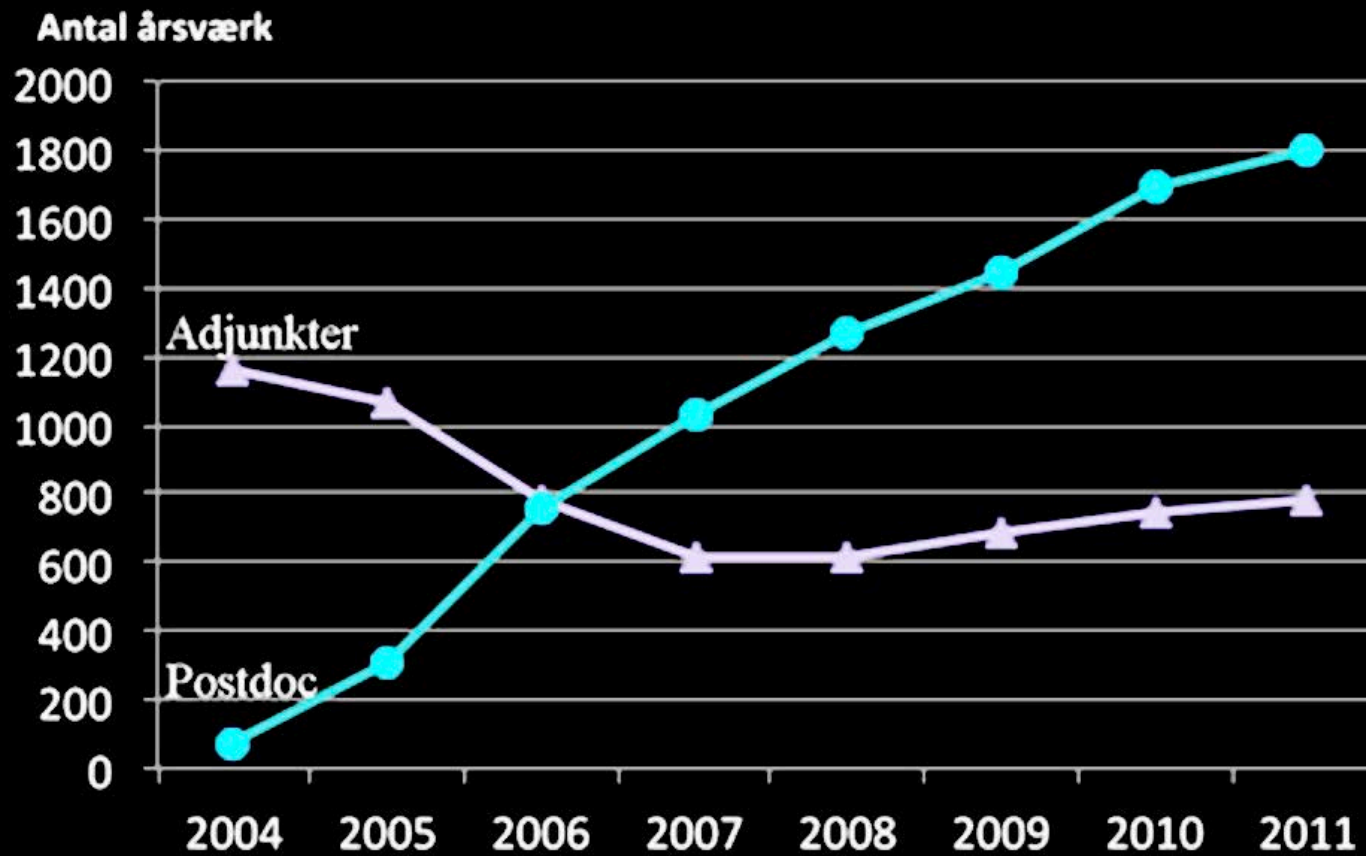
White et al.
(2012)

Citations-prediktion

Tanner et al.
(2015)

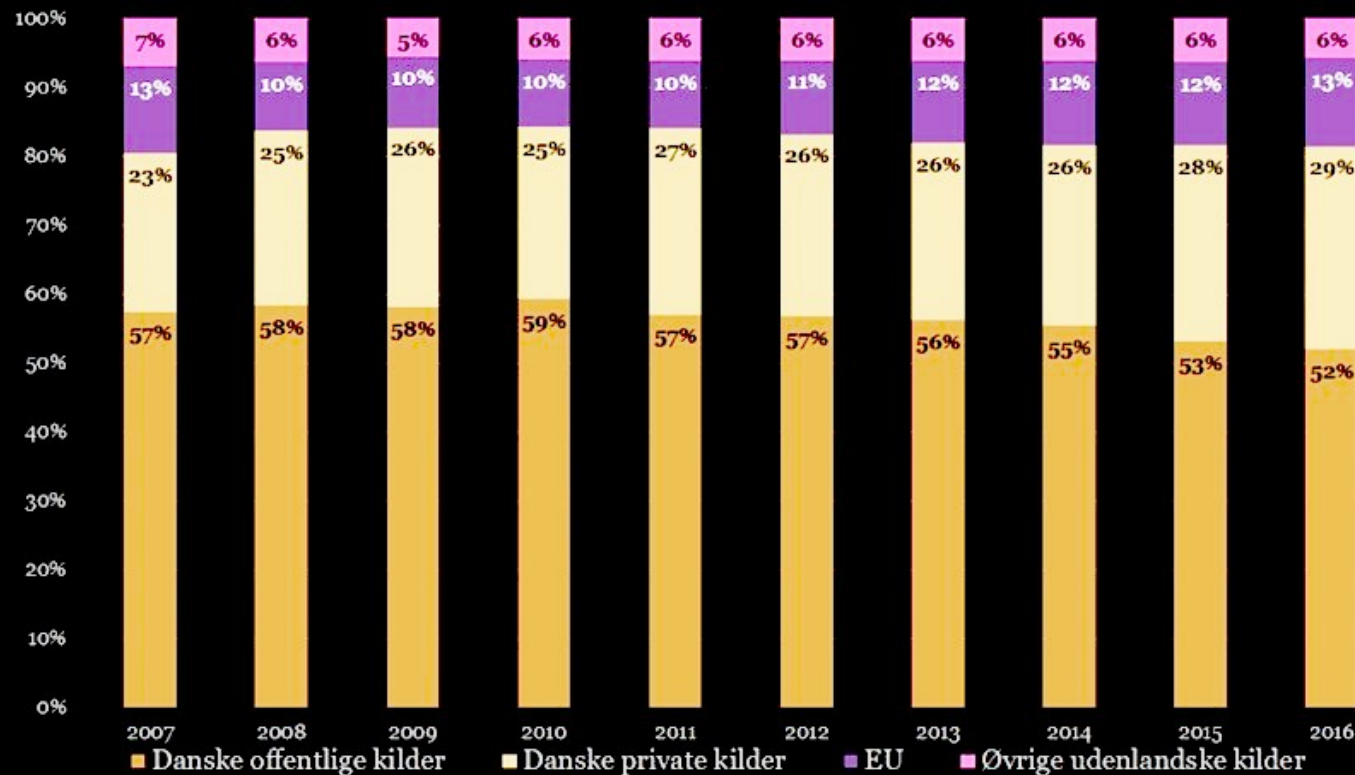
Nye krav til
forskningssupport

- Øget vigtighed
- Overblik over funding-muligheder (virksomheder, crowdfunding, fonde)
- Dokumentation af impact



- Øget vigtighed
- Overblik over funding-muligheder (virksomheder, crowdfunding, fonde)
- Dokumentation af impact

Fordelingen af eksterne kilder til finansiering af forskning på universiteter



- Øget vigtighed
- Overblik over funding-muligheder (virksomheder, crowdfunding, fonde)
- Dokumentation af impact



Målbar impact?

Metrikker

Data-drevet

Virksomheds-funding

Indflydelsesrige citationer (Semantic Scholar)

Pressedækning

Alumni

Påvirkning af diskurs og opinion

Forudsigelse: Ingen rapportering

Forslag: Lad forskere vælge deres metrik

Forudsigelse af impact

Forudsigelse af manuskript-success

Industry

Forudsigelse af virale nyheder

Industry

Forudsigelse af svars relevans

Johannsen og
Søgaard (2013)

Detektion af videnskabelig snyd

Braud og
Søgaard (2018)

Automatiseret peer review

Nguyen et al.
(2016)

Forudsigelse af impact

Forudsigelse af manuskript-success

Industry

Forudsigelse af virale nyheder

Industry

Forudsigelse af svars relevans

Johannsen og
Søgaard (2013)

Detektion af videnskabelig snyd

Braud og
Søgaard (2018)

Automatiseret peer review

Nguyen et al.
(2016)

Forudsigelse af impact

Yogatama et al.
(2011)

Predicting a Scientific Community's Response to an Article

Dani Yogatama Michael Heilman Brendan O'Connor Chris Dyer

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA

{dyogatama, mheilman, brenocon, cdyer}@cs.cmu.edu

Bryan R. Routledge

Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213, USA
routledge@cmu.edu

Noah A. Smith

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
nasmith@cs.cmu.edu

Abstract

We consider the problem of predicting measurable responses to scientific articles based primarily on their text content. Specifically, we consider papers in two fields (economics and computational linguistics) and make predictions about downloads and within-community citations. Our approach is based on generalized linear models, allowing interpretability; a novel extension that captures first-order temporal effects is also presented. We demonstrate that text features significantly improve accuracy of predictions over metadata features like authors, topical categories, and publication venues.

marization, we consider supervised models of the collective *response* of a scientific community to a published article. There are many measures of impact of a scientific paper; ours come from direct measurements of the number of downloads (from an established website where prominent economists post papers before formal publication) and citations (within a fixed scientific community). We adopt a discriminative approach based on generalized linear models that can make use of any text or metadata features, and show that simple lexical features offer substantial power in modeling out-of-sample response and in *forecasting* response for future articles. Realistic forecasting evaluations require methodological care beyond the usual best practices of train/test separation, and we elucidate these issues.

1 Introduction



Resumé

- Mindre offentlig støtte, men flere gaver
- Flere steder fra
- Større, men mere diverst, impact-krav
- **Automatiseret discovery**
- Automatiseret impact estimation
- Bonus: Øv virksomhedsorienteret sales-pitch



Innovationsfonden

facebook®

CARISBERG FOUNDATION



Joachim Bingel

Postdoc
NLP, Dialogue, Simplification



Ana Valeria Gonzalez

PHD Student
NLP, Dialogue



Mareike Hartmann

PHD Student
NLP, Social Science



Desmond Elliott

Assistant Professor
NLP, Dialogue, Simplification



Victor Petrn Hansen

PHD Student
NLP, Dialogue



Maria Barrett

Postdoc
NLP, Dialogue, Simplification



Yova Kementchedjheva

PHD Student
NLP, Unsupervised Learning



Mostafa Abdou

PHD Student
NLU, Semantic Parsing



Anders Søgaard

Professor
NLP, Machine Learning



Simon Flachs

PHD Student
NLP, Error correction



Rahul Aralikkatte

PHD Student
NLU, Machine Learning



Daniel Hershovich

Postdoc
NLP



Heather Lent

PHD
NLP



Lasse Borgholt

PHD
NLP, Speech



Marcel Bollmann

Postdoc
NLP, Morphology





Resumé

- Mindre offentlig støtte, men flere gaver
- Flere steder fra
- Større, men mere diverst, impact-krav
- **Automatiseret discovery**
- Automatiseret impact estimation
- Bonus: Øv virksomhedsorienteret sales-pitch



Innovationsfonden

facebook®

CARISBERG FOUNDATION



Joachim Bingel

Postdoc
NLP, Dialogue, Simplification



Ana Valeria Gonzalez

PHD Student
NLP, Dialogue



Mareike Hartmann

PHD Student
NLP, Social Science



Desmond Elliott

Assistant Professor
NLP, Dialogue, Simplification



Victor Petrn Hansen

PHD Student
NLP, Dialogue



Maria Barrett

Postdoc
NLP, Dialogue, Simplification



Yova Kementchedjheva

PHD Student
NLP, Unsupervised Learning



Mostafa Abdou

PHD Student
NLU, Semantic Parsing



Anders Søgaard

Professor
NLP, Machine Learning



Simon Flachs

PHD Student
NLP, Error correction



Rahul Aralikkatte

PHD Student
NLU, Machine Learning



Daniel Hershovich

Postdoc
NLP



Heather Lent

PHD
NLP



Lasse Borgholt

PHD
NLP, Speech



Marcel Bollmann

Postdoc
NLP, Morphology



?

coASTal



UNIVERSITY OF
COPENHAGEN