Dutch Humor Detection by Generating Negative Examples

Thomas Winters & Pieter Delobelle

PhD Students at DTAI, KU Leuven

firstname.lastname@kuleuven.be

@thomas_wint

thomaswinters.be

@pieterdelobelle

people.cs.kuleuven.be /~pieter.delobelle

Humor

Intrinsically human!

AI-Complete problem?

Incongruity-Resolution Theory

Setup

Two fish are in a tank. Says one to the other:

Punchline

"Do you know how to drive this thing?"

Obvious Interpretation



Hidden Interpretation



Based on: Ritchie, G. (1999). Developing the incongruity-resolution theory.

Human-focused definition!

Machine should **not only spot two mental images**

Obvious Interpretation



Hidden Interpretation



But also this is **not too hard or too easy** for a **human**!

Transformer models

Large **language models**, **pretrained** on large corpora

OpenAI



GPT-2 & GPT-3

Completes any textual prompt

BERT

Classifies any text sequence / token

Outperforming previous neural architectures
on **most language tasks**

Brown, Tom B., et al. "Language models are few-shot learners."
Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding

Not just for English → Dutch RobBERT!



RobBERT

Our Dutch BERT-like model

RobBERT is a **Dutch RoBERTa**-based language model

Vastly outperforms other architectures on **large range of Dutch NLP tasks** & generally outperforms other BERT models, especially on **small datasets**

Easy to use: just import & fine-tune on your task

```
from transformers import RobertaTokenizer, RobertaForSequenceClassification
tokenizer = RobertaTokenizer.from_pretrained("pdelobelle/robbert-v2-dutch-base")
model = RobertaForSequenceClassification.from_pretrained("pdelobelle/robbert-v2-dutch-base")
```

But can it **learn to recognise humor**?

Early Humor Detector

- Designed **humor features** *e.g. alliteration, antonym, adult slang…*
- Used **Naive Bayes** and **Support Vector Machines**
- **Task:** One-liners vs **news**, **neutral corpus** & **proverbs**

One-liners Reuters	One-liners BNC	One-liners Proverbs
96.95%	79.15%	84.82%

Table 4: Humor-recognition accuracy for combined learning based on stylistic and content features.

Mihalcea, R., & Strapparava, C. (2005). Making computers laugh: Investigations in automatic humor recognition.

But is this a good dataset?

News & proverbs have **completely different types** of words than jokes!



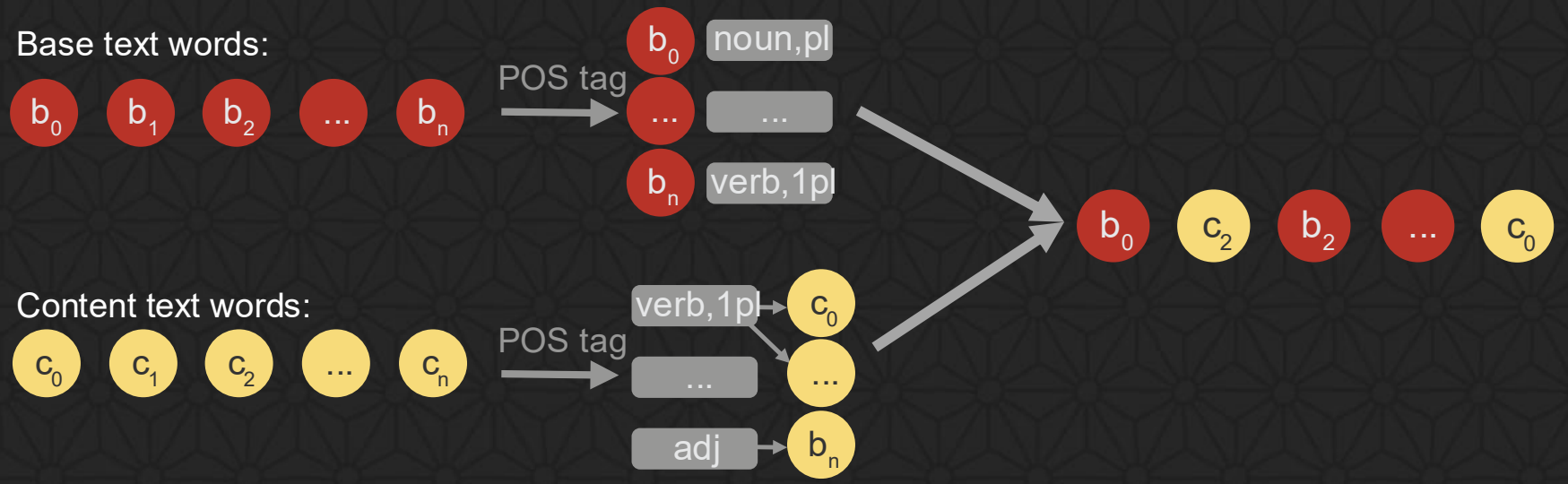


→ Looking at **word frequencies** is often already *"enough"*!

Is this really **humor** detection?

Jokes are **fragile**!

Two men are in a bar. Says one to the other:
"Do you know how to drive this thing?"



Generate non-jokes using **dynamic templates**! *(@TorfsBot)*

Word-based features **won't work** anymore!

Examples of generated Dutch non-jokes

Het is groen en het is een mummie?
Kermit de Waterkant

Wat is het toppunt van principe?
1) Wachten totdat een Nederlander gaat twijfelen
2) Een Zuster met een autoladder
3) Een brandwacht brandmeester met een brandmeester van 9 maanden

“Ober, kunt u die schrik uit mijn politieman halen? Want ik eet liever alleen.”

"Mijn hond is heel vreselijk: Hij schreeuwt mij iedere zus de broer.“
"Maar dat is toch niet zo heel vreselijk?“
"Jawel, want ik heb geen rapport!"

Wat staat er midden in het bos?
De kapper.

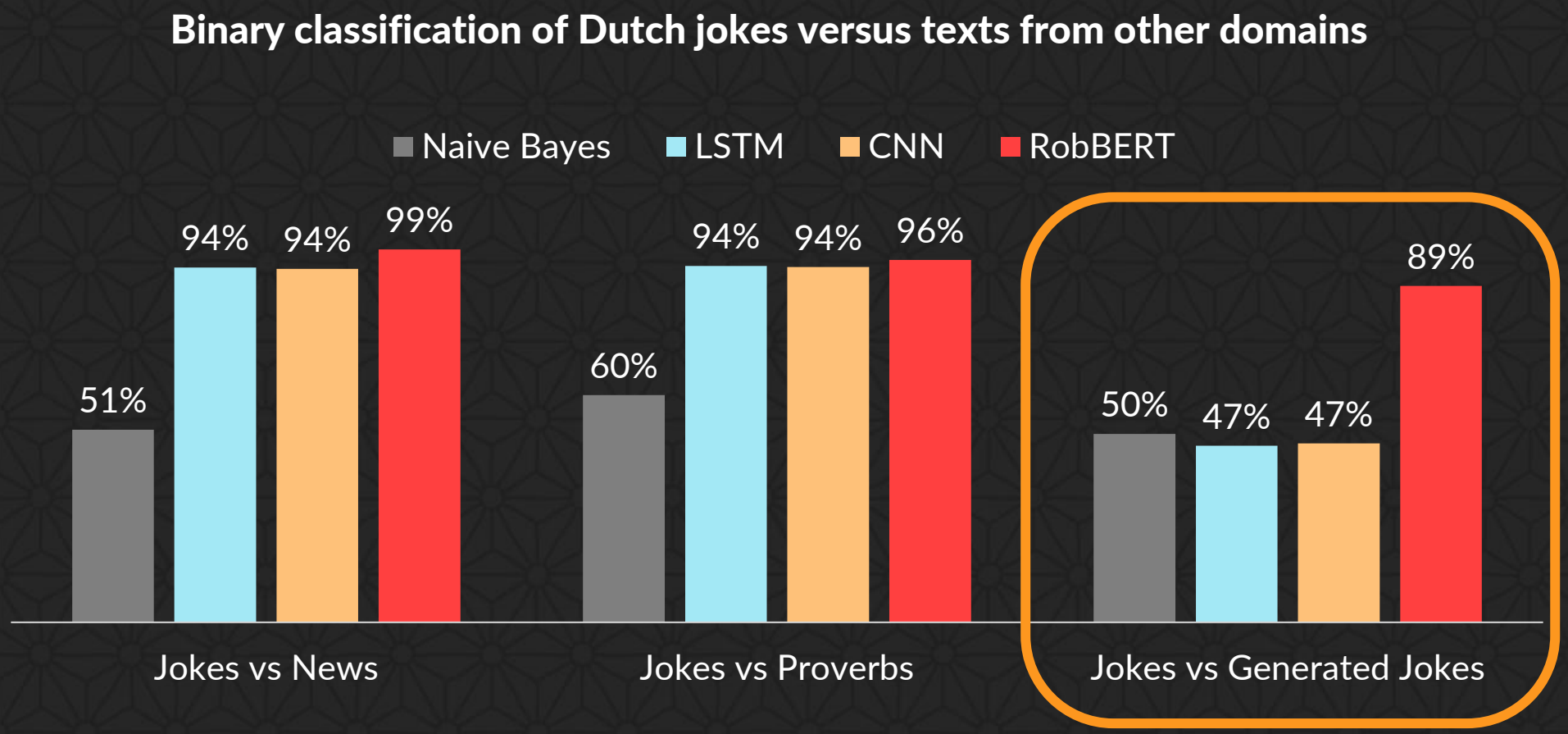
Er loopt een super vriendelijk blondje langs een armband. Last er een toonbank: “zo, waargaan die mooie mannen heen?” Blondje: “naar de barkeeper als er niets tussen komt…”

Hoe heet de vrouw van Sinterklaas?
Keukentafel.

"Twee tanden zwemmen in de zee en ze zien een stamgast op een stamgast. De ene raad zegt tegen de andere raad: 'Hé kijk! Ons eten op een bord!'"



Much more challenging dataset!



More truthful humor detection?

Conclusie



Novel joke detection dataset creation method

Easily scales to other languages



Illustrated humor insights of transformer

Strongly outperforming previous neural networks



Created first Dutch humor detectors

https://github.com/twinters/dutch-humor-detection



Dutch Humor Detection by Generating Negative Examples

Thomas Winters & Pieter Delobelle

PhD Students at DTAI, KU Leuven

firstname.lastname@kuleuven.be

@thomas_wint

thomaswinters.be

@pieterdelobelle

people.cs.kuleuven.be/~pieter.delobelle

Some images (based on the works) of dooder on freepik.com