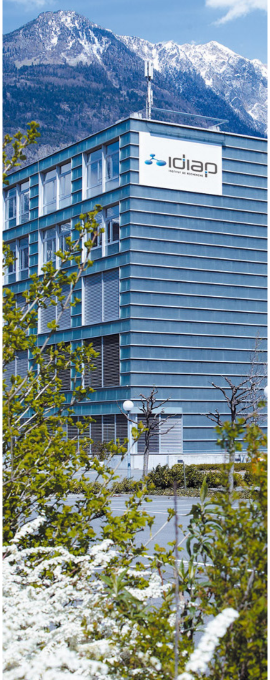


Automatic Diagnosis of Alzheimer's Disease Using Neural Network Language Models

Julian Fritsch, Sebastian Wankerl, Elmar Nöth

May 17, 2019

idiap RESEARCH INSTITUTE

EPFL

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Outline

1. Problem Statement
2. Language Modeling based Alzheimer's Classification
3. Experimental Methodology
4. Results & Analysis
5. Conclusions

How to Assess a Demented Person’s Cognitive State

- Alzheimer’s dementia is a neurodegenerative disease

Mini-mental state exam (MMSE)

- executed by a physician
- 30 questions to assess mental capabilities:
  - score < 19: severe dementia
  - score > 29: median of healthy people

How to Assess a Demented Person's Cognitive State

- Alzheimer's dementia is a neurodegenerative disease

Mini-mental state exam (MMSE)

- executed by a physician
- 30 questions to assess mental capabilities:
  - score < 19: severe dementia
  - score > 29: median of healthy people

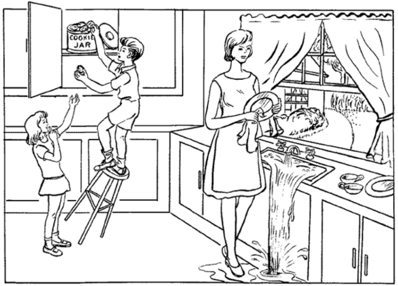
Automatic analysis of spontaneous speech

- **Cookie Theft picture description**

Alzheimer's Classification based on Language Structures

Cookie Theft picture description

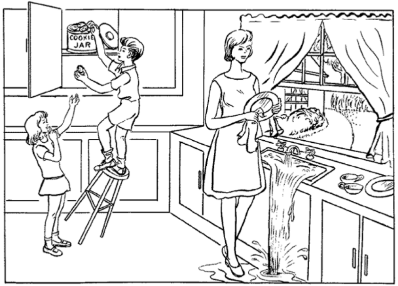
- natural approximation to spontaneous discourse



Alzheimer's Classification based on Language Structures

Cookie Theft picture description

- natural approximation to spontaneous discourse



Alzheimer's patient:

- « There's a young boy getting a cookie jar.
  And it he's uh in a bad shape because uh the thing is falling over. »

Healthy control:

- « A boy is trying to get cookies out of a jar
  and he's about to tip over on a stool. »

Language Modeling based Alzheimer's Classification

Language modeling

- assigning probabilities $P(w)$ to words given previous words

« the high ... tree/tower/mountain »

Language Modeling based Alzheimer's Classification

Language modeling

- assigning probabilities $P(w)$ to words given previous words

« the high ... tree/tower/mountain »

Language model evaluation: perplexity (PPL)

- $PPL(S) = P(S)^{-\frac{1}{N}}$, $P(S)$:= probability of sequence S, N := # words in S

Language Modeling based Alzheimer's Classification

Language modeling

- assigning probabilities $P(w)$ to words given previous words

« the high ... tree/tower/mountain »

Language model evaluation: perplexity (PPL)

- $PPL(S) = P(S)^{-\frac{1}{N}}$, $P(S)$:= probability of sequence S, N := # words in S

Perplexity-based Alzheimer's classification using n-grams [1]

- perplexity difference used for binary Alzheimer's classification
- n-grams have a fixed context length

[1] S. Wankerl, E. Nöth, and S. Evert, "An n-gram based approach to the automatic diagnosis of alzheimer's disease from spoken language," in Proc. Interspeech, 2017.

Experimental Methodology

Address shortcomings of n-grams: RWTHLM toolkit

- building and evaluating neural network language models (NNLMs)
- designed for using recurrent and long short-term memory (LSTM) layers
  $\rightarrow$ allowing variable context length

Experimental Methodology

Address shortcomings of n-grams: RWTHLM toolkit

- building and evaluating neural network language models (NNLMs)
- designed for using recurrent and long short-term memory (LSTM) layers
  $\rightarrow$ allowing variable context length

Experimental setup

- LMs from Alzheimer's $\mathcal{M}_{Alzheimer's}$ and control transcriptions $\mathcal{M}_{control}$
- leave-one-speaker-out cross-validation
- excluding 10 randomly selected speakers for validation

Experimental Methodology

Address shortcomings of n-grams: RWTHLM toolkit

- building and evaluating neural network language models (NNLMs)
- designed for using recurrent and long short-term memory (LSTM) layers
  $\rightarrow$ allowing variable context length

Experimental setup

- LMs from Alzheimer's $\mathcal{M}_{Alzheimer's}$ and control transcriptions $\mathcal{M}_{control}$
- leave-one-speaker-out cross-validation
- excluding 10 randomly selected speakers for validation

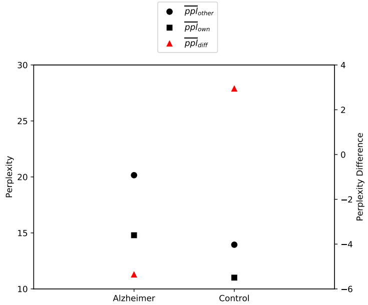
Evaluation

- perplexity evaluation of each speaker s on 2 LMs giving ppl_{own} and ppl_{other}

$$ppl_{diff} = \begin{cases} ppl_{own} - ppl_{other} & \text{if } s \in \text{Alzheimer's group} \\ ppl_{other} - ppl_{own} & \text{if } s \in \text{control group} \end{cases}$$

Perplexity Difference for Binary Classification

Comparison of perplexity means from both groups



- classification threshold at equal-error rate (EER)

Data – DementiaBank's Pitt Corpus

English Cookie Theft picture descriptions & MMSE scores

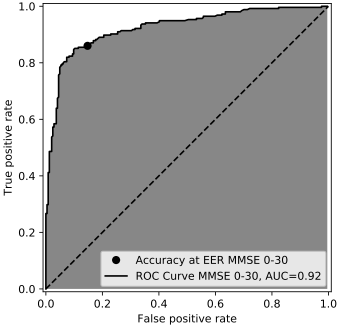
- conducted yearly
- publicly available

Selection for Alzheimer's classification:

- 168 Alzheimer's patients, 255 transliterations
- 98 control patients, 244 transliterations

Performance Evaluation with ROC Curves (1)

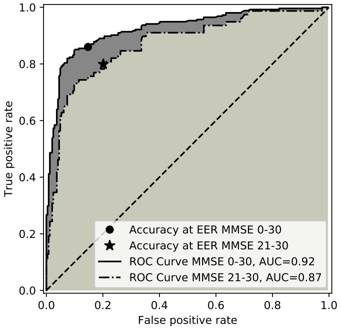
Overall accuracy: **85.6%** at EER, 72 wrongly classified transliterations (compared to 77.1% at EER with tri-grams)



Performance Evaluation with ROC Curves (2)

Overall accuracy: **85.6%** at EER, 72 wrongly classified transliterations (compared to 77.1% at EER with tri-grams)

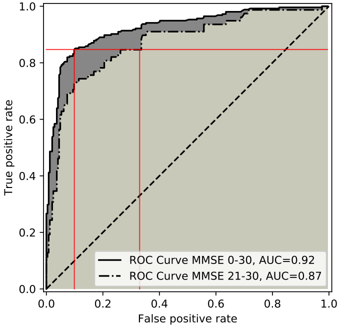
Speakers with an MMSE score from 21 to 30: **79.9%** at EER, 66 wrongly classified transliterations



Performance Evaluation with ROC Curves (3)

All speakers: 85% true positive rate (TPR), 10% false positive rate (FPR)

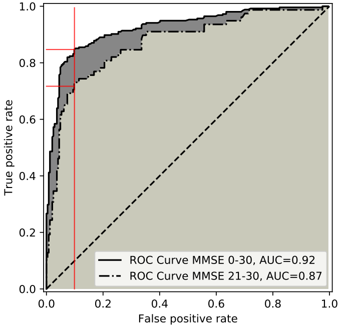
Speakers with an MMSE score from 21 to 30: 85% TPR, 33% FPR



Performance Evaluation with ROC Curves (4)

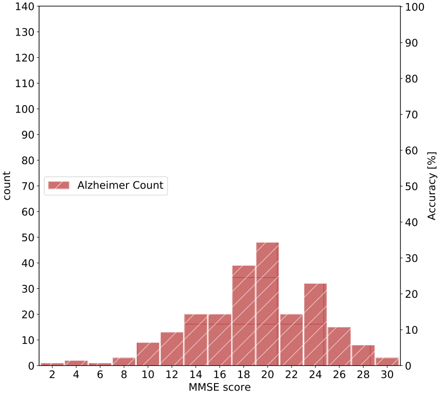
All speakers: 85% true positive rate (TPR), 10% false positive rate (FPR)

Speakers with an MMSE score from 21 to 30: 73% TPR, 10% FPR



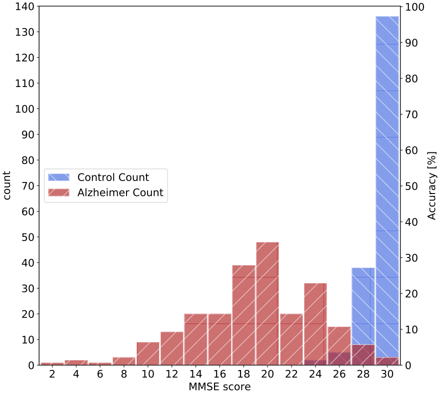
Classification Results per MMSE (1)

Histogram of all Alzheimer's MMSE scores



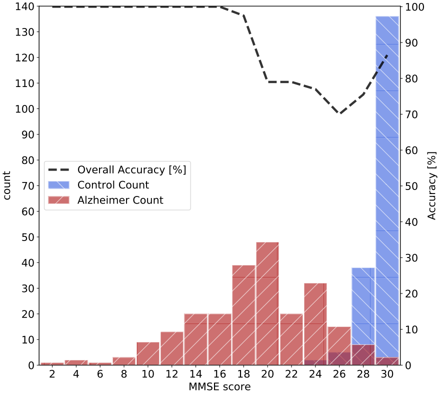
Classification Results per MMSE (2)

Histogram of all Alzheimer's and control MMSE scores



Classification Results per MMSE (3)

Histogram of all Alzheimer's and control MMSE scores and accuracy per MMSE



Using Perplexity Difference for MMSE Estimation

Pearson's correlation r and Spearman's correlation ρ between MMSE scores and perplexity difference p_{diff}:

	r	ρ
Alzheimer's	0.433	0.547
Control	0.112	0.109
All	0.656	**0.771**

Conclusions

Neural network-based language models used for Alzheimer's classification

- model language structures well (**85.6%** vs 77.1% with tri-grams)
- perplexity difference correlates well with MMSE scores
- is a purely statistical approach transferable to other languages

Conclusions

Neural network-based language models used for Alzheimer's classification

- model language structures well (**85.6%** vs 77.1% with tri-grams)
- perplexity difference correlates well with MMSE scores
- is a purely statistical approach transferable to other languages

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie grant agreement No 766287.

ROC Curve Comparison to N-grams

ROC Curve Comparison to N-grams

LSTM-NNMLs: **85.6%** at EER, 72 wrongly classified transliterations

Tri-gram LMs: **77.1%** at EER, 114 wrongly classified transliterations