Fall Detection in EHR using Word Embeddings and Deep Learning

Henrique Santos - PUCRS - Brazil
19th IEEE International Conference on Bioinformatics and Bioengineering

Brief context …

Multidisciplinary Group and ...

PUCRS
Pontifical Catholic University - Rio Grande do Sul

PPGCC and PPGGB
Graduate program in Computer Science
Graduate program in Biomedical Gerontology

… and Multicentric Group

Research Group on **Artificial Intelligence** in Healthcare

GHC
Grupo Hospitalar Conceição

HOSPITAL
MÃE DE DEUS
SISTEMA DE SAÚDE MÃE DE DEUS

HOSPITAL
MOINHOS
DE VENTO

SANTA CASA
DE MISERICÓRDIA
PORTO ALEGRE

SOCIEDADE BENEFICENTE DE SENHORAS
SÍRIO·LIBANÊS

HOSPITAL
SÃO LUCAS
DA PUCRS

http://www.inf.pucrs.br/ia-saude/

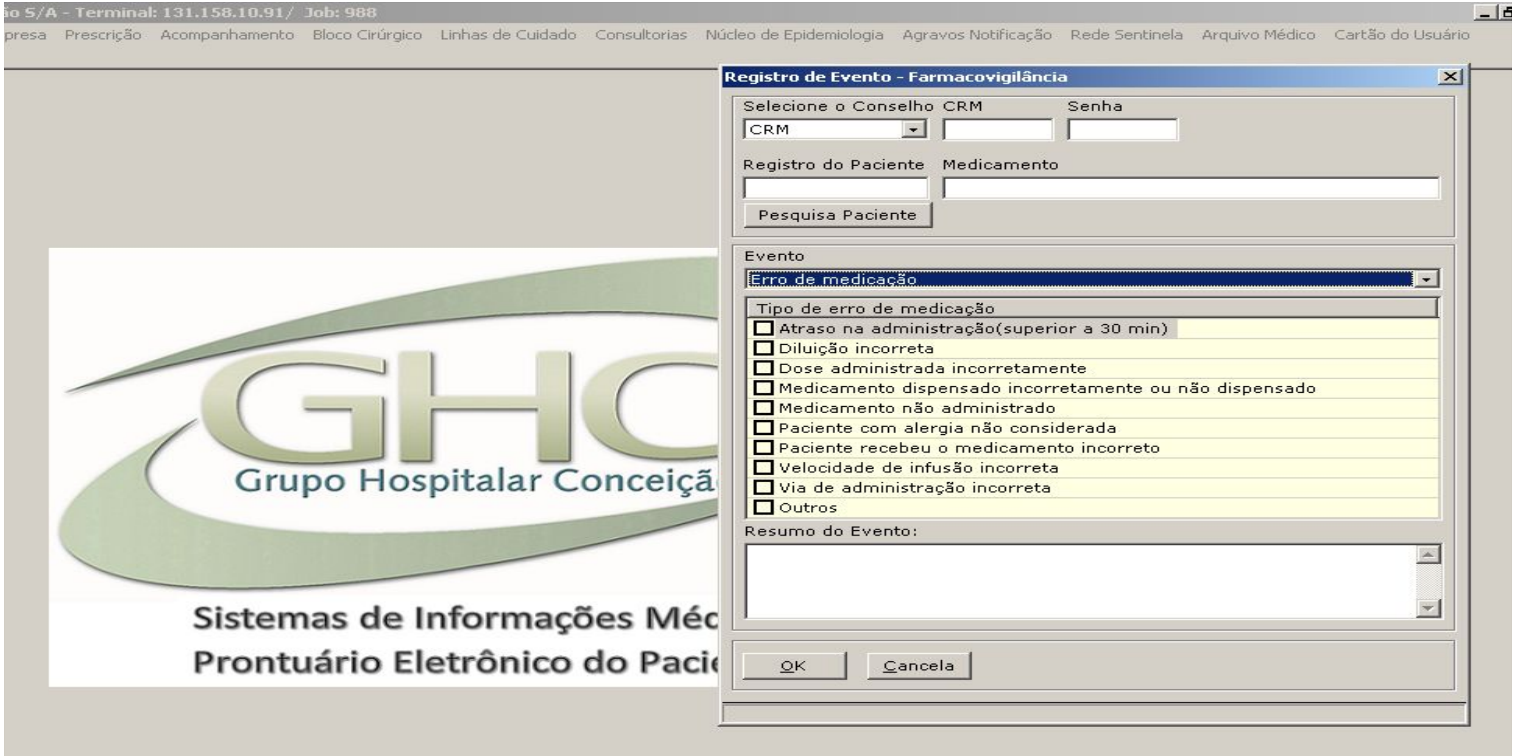
Scope of this Research

Fall

This adverse event is very common in the hospital environment. The Morse Fall Scale measure the risk of patient fall.

There is a underreport number of fall event notifications.

Report System



Underreported Events

Events are notified voluntarily in the system

Source of useful information to improve care and subsidize continuing education!

Due to lack of knowledge, forgetfulness or lack of time, it is underreported. Only 10-20% of these events are reported.

IHI Global Trigger Tool

Institute *for* Healthcare Improvement

"This tool includes a list of known AE triggers as well as **instructions for selecting records**, training information, and appendices with references and common questions."

Problems and Objectives

Detect falls incidents in Progress Notes

Evaluate language models performance using LSTM in this task

Methods and Materials

Word Embeddings

Variations

surgery , surgeri , srg

surge , surger , urgery ...

Relations

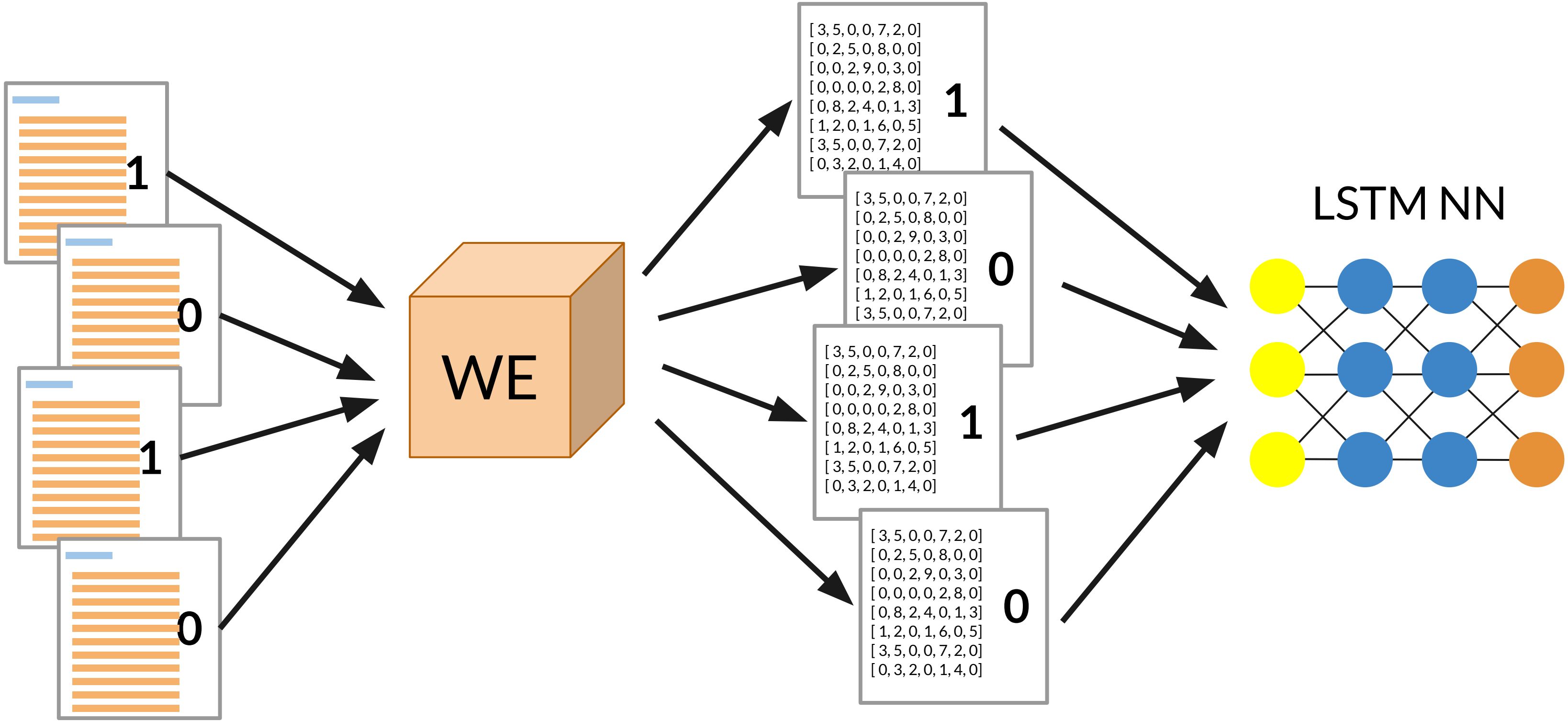
acetaminophen → fever

amoxicillin → infection

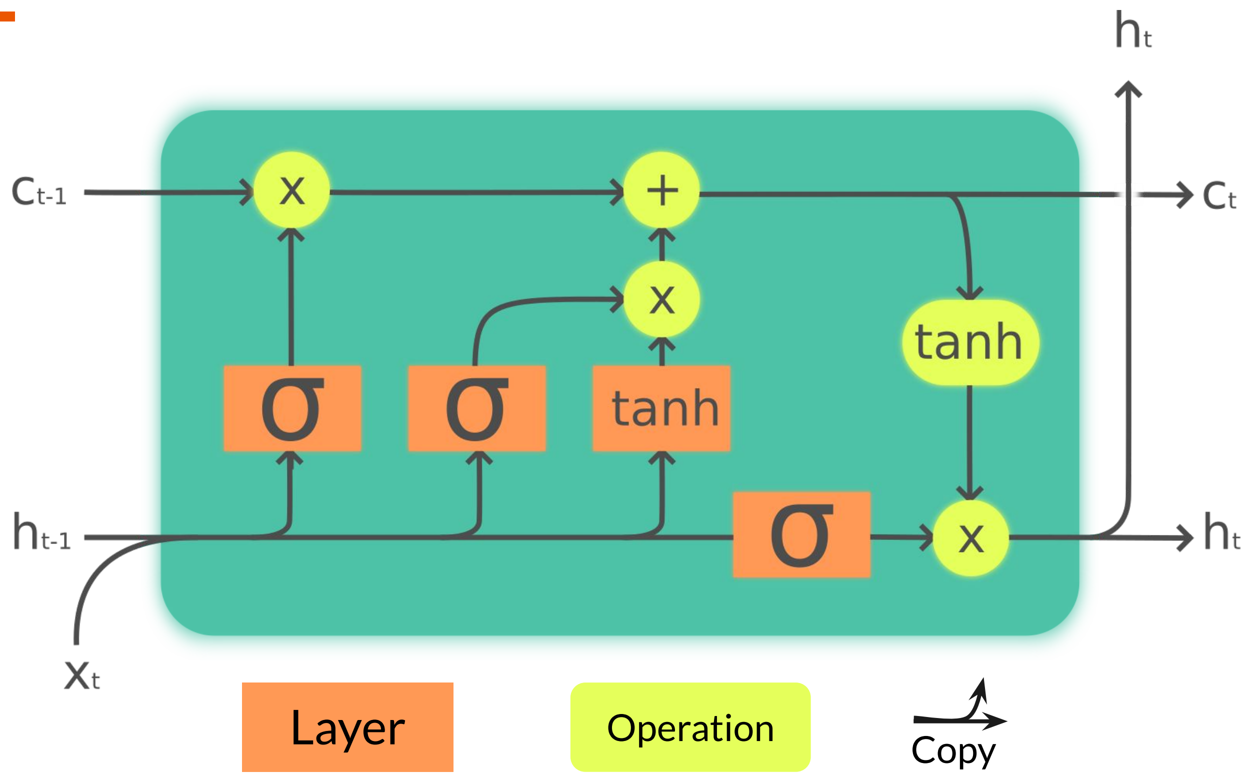
amlodipine → vasodilator

neural network unsupervised learning

Word Embeddings Encoding



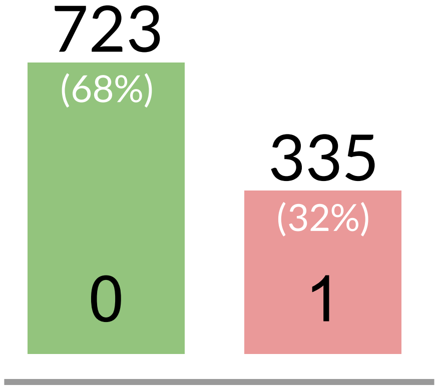
LSTM Neuron (Long Short-Term Memory)



Annotation Process

441
Incident Reports

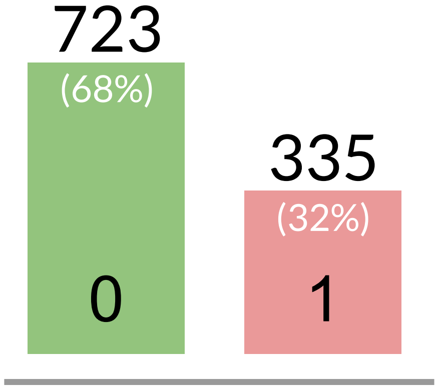
Annotation Process

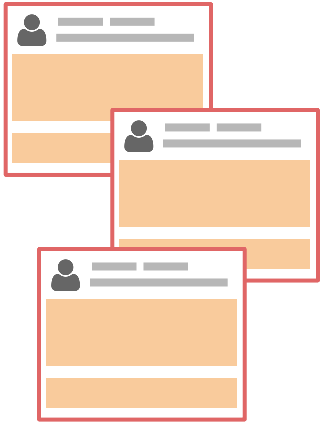


441
Incident Reports

1,078
Progress Notes

Annotation Process





441
Incident Reports

1,078
Progress Notes

367
Patients

Annotation Process

Fall per Patient in Annotated Dataset

# of Patients	% of Total	# of Falls
316	87.0 %	1 fall
36	10.0%	2 falls
11	3.0 %	3 falls
1	0.3 %	4 falls
...	...	...

Language Models

Wikipedia-PT

- Wikipedia dump of May 2019
- 250 million tokens

NILC (broad-domain)

- Wikipedia + 20 sources (News and Researches)
- 1.3 billion tokens

EHR-Notes (biomedical-domain)

- 24 million sentences from Progress Notes
- 603 million tokens

Fall Detection - Experiments

Annotation of 1,078 Progress Notes

Evaluation of three Word Embedding Models

- Wikipedia-PT
- NILC (Wikipedia + 20 sources)
- EHR-Notes (Progress Notes)

State-of-the-art NLP Neural Network (LSTM)

Baseline: SVM and Random Forest

Results ...

Fall Detection - Results

Embeddings 5-Fold Cross Validation: F-Measure

	Word2Vec	FastText
Wikipedia	0.88 ± 0.14	0.87 ± 0.11
NILC (broad-domain)	0.77 ± 0.06	0.89 ± 0.13
EHR (biomedical-domain)	0.88 ± 0.14	**0.90 ± 0.13**

Fall Detection - Results

Embeddings 5-Fold Cross Validation: F-Measure

	Word2Vec	FastText
Wikipedia	0.88 ± 0.14	0.87 ± 0.11
NILC (broad-domain)	0.77 ± 0.06	0.89 ± 0.13
EHR (biomedical-domain)	0.88 ± 0.14	**0.90 ± 0.13**
Random Forest	0.73 ± 0.03	
SVM	0.60 ± 0.05	

Conclusion

Better with biomedical-domain Language Model

LSTM prove to be best than Random Forest

Research Limitations

- Progress Notes Sampling (natural distribution)
- Self-Attention Neural Network
- BERT, GPT-2, XLNet Embeddings
- Quality Evaluation (why right, why wrong)

Source Code

- Experiments scripts
- Annotated Dataset (1,078 records)
- Pre-trained Word Embeddings

https://github.com/nlp-pucrs/fall-detection

Further Work ...

Further Work

Sequence Tagging Task : F-Measure

	LSTM + CRF
Wikipedia	0.55
NILC (broad-domain)	0.67
EHR (specific-domain)	**0.80**

Further Work

Quality Embeddings - Analogies Task

	Specific	**Broad**
Wikipedia	1.38 %	79.00 %
NILC (broad-domain)	2.61 %	**82.38 %**
EHR (specific-domain)	**2.85 %**	0.00 %

Thanks! Questions?

http://www.inf.pucrs.br/ia-saude/

Henrique Dias

henrique.santos.003@acad.pucrs.br

Annotation Process - WebAnno