Deep Attention Model for Triage of Emergency Department Patients

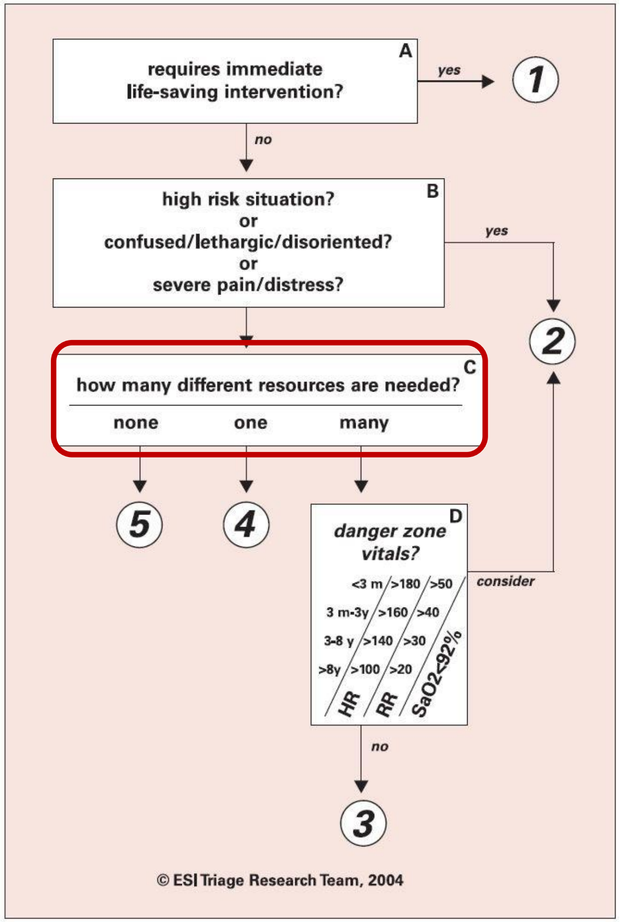
Djordje Gligorijevic
gligorijevic@temple.edu

Gligorijevic, Dj., Stojanovic, J., Wayne, S., Stojkovic, I., Schreyer, K., Del Portal, D., Obradovic, Z., (2018) "Deep Attention Model for Triage of Emergency Department Patients", Proceedings of the 2018 SIAM International Conference on Data Mining (SDM 2018), San Diego, CA, May 2018

Triage Process in Emergency Departments

- **Existing solution**: Emergency Severity Index (ESI) system for patient triage was introduced to help guide manual estimation of acuity levels, which is used by nurses to rank the patients and organize hospital resources:

Principal issue: human variability in the equation.
ESI: 5 levels of acuity
Number of resources categories: 0-5 (6 of them)
Note: they are not the same, but are correlated

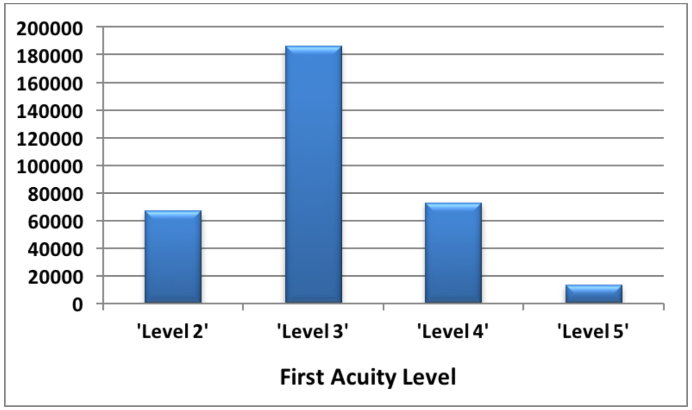


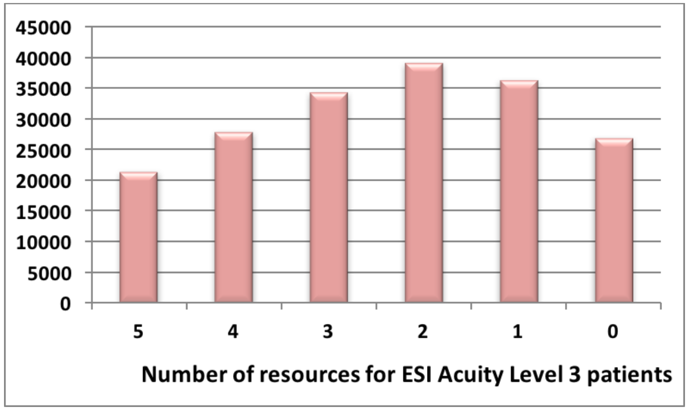
Emergency Severity Index (ESI): A Triage Tool for Emergency Department

Deep Attention Model for Triage of Emergency Department Patients: Problem and Challenge

Motivation: Current solutions have consistency problems and bring many uncertainties yielding suboptimal triage process

- most of patients are assigned ESI level 3 while in fact they are not
- patient are processed in the first come-first serve basis





Challenge: Use nurses' notes data to predict ED visits outcome for an automatic and accurate triage.

Predictive tasks: i) **Patients severity** (binary classification : $\geq$ 3 resources is severe),
ii) **Resources needed** (multi-class classification: 6 resource classes)

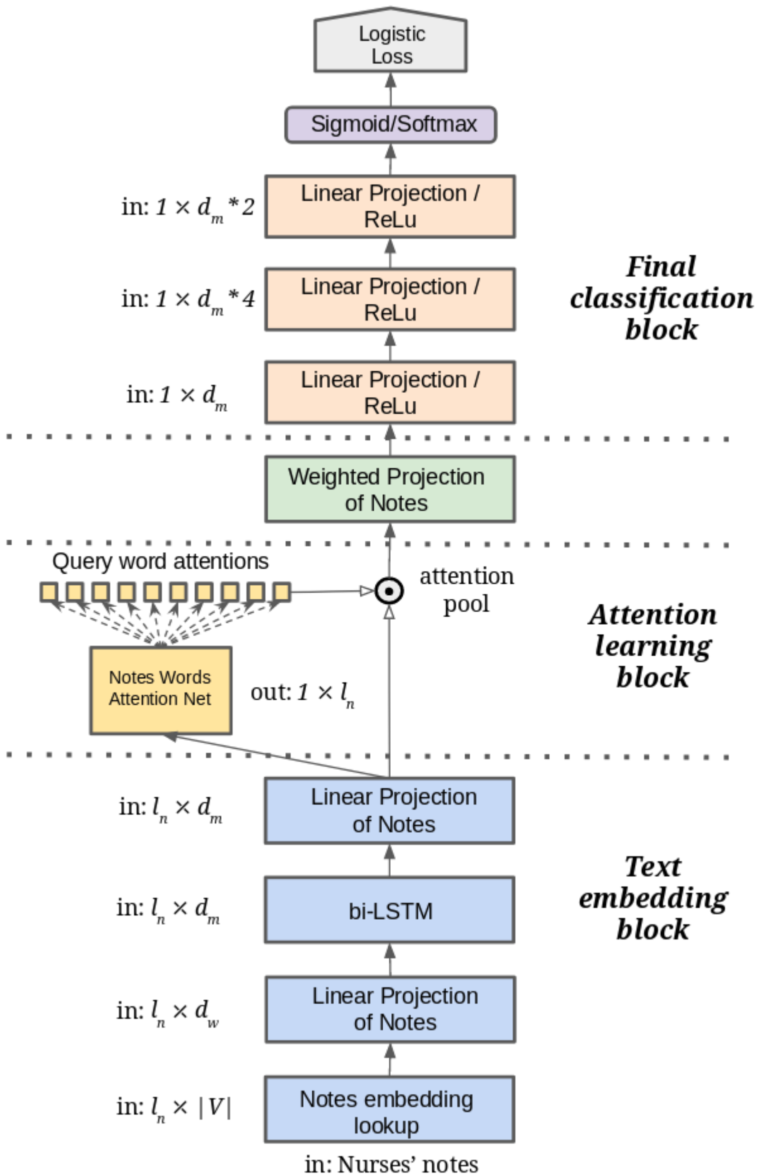
Deep Attention Model for Triage of Emergency Department Patients: Data

- **Source**: Retrospective review of Temple University ED data over a period 2012.– 2015.

- **Data elements**:
  - i. **Structured** (continuous and nominal) data **(p_s)** includes:
    - i. ED assigned location,
    - ii. gender,
    - iii. age range,
    - iv. method of arrival,
    - v. hour of arrival,
    - vi. number of prior ED visits,
    - vii. insurance group,
    - viii. heart rate,
    - ix. systolic blood pressure, and
    - x. temperature.
  - ii. **Unstructured** (text) data **(p_u)** includes:
    - i. chief complaint,
    - ii. past medical history,
    - iii. medication list, and
    - iv. free text initial nursing assessment.
  - iii. The response variable represents category of number of resources (0-5)

- The **data** consists of 338,500 ED visits
- The **test** set is comprised of 68,500 patients out of which 36,883 are ESI level 3

Deep Attention Model (DAM) for Triage of Emergency Department Patients: The model



Classification block is used to classify patient into one of the 6 categories. Additionally, in this layer we add handcrafted features from measurements taken during the triage

Attention block is used to add scores to words in the document to allow the model to focus on the most important words and phrases

In **text embedding block** words in the nurses notes are embedded in the low dimensional space using multiple linear layers with bi-directional LSTM layer to learn higher order relations among them.

Learn nurses’ notes summarization using attention mechanism

- **Learning nurses’ notes summarization:** We propose attention block is to learn scores of words in the document used to summarize given word sequence

- Other summarization strategies: Max vector, **Sum vector** (best performing), Average vector, Product vector.
- Sum vector representation:

$$v_n = \sum_i h_n^{(i)}$$

- Attention layer $s\left(h_n^{(i)};\ \Theta\right)$ is a separate neural network architecture that learns to score words based on their embeddings using softmax function:

$$a(i) = \frac{\exp(s\left(h_n^{(i)};\Theta\right))}{\sum_{i=1}^{l_n} s\left(h_n^{(i)};\Theta\right)}.$$

- two fully connected layers with ReLU nonlinearities are used in our experiments

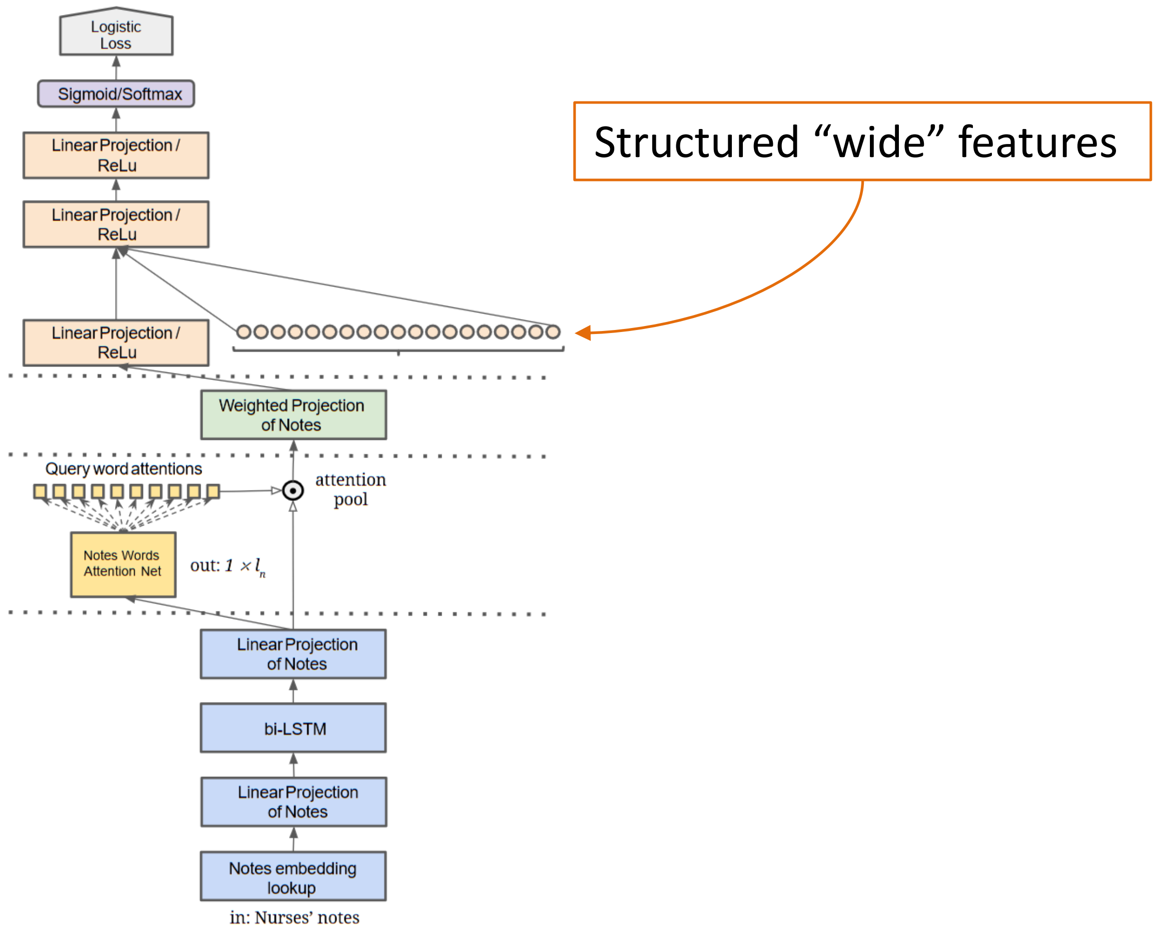
- Learning attentions is coupled with the entire network in our model

- The final vector projection of nurses’ notes is then obtained as:

$$v_n = \sum_i a^{(i)} * h_n^{(i)}$$

Deep & Wide architecture

- Use structured data in the "wide" part of the model:



- Obtain **Deep&Wide architecture by** appending structured features to the learned text representation

Experimental baselines

- In our experiments we employ models that use:

  - Only structured data:
    1. Logistic Regression (**LogReg**)
    2. Multi-layered Perceptron (**MLP**)

  - Both unstructured and structured data:
    3. Bi-directional LSTM model (**bi-LSTM**)
    4. word-level Very Deep CNN model (**VDCNN**)
    5. Deep sum-pooling model (**DSMP**) – variant of DAM with sum-pool instead of attention mechanism

All deep models are evaluated with and without using handcrafted features to investigate whether using them provides lift in accuracy.

Research Question 1: Can free text data be utilized? Does adding structured features help?

- **Q**: Can we automatically learn useful representations from the nurses' notes textual content, without any feature handcrafting
- **Q**: Are additional **structured features** helpful for this task?

	Binary		Multi-class	
Models	**Acc.**	**AUC**	**Acc.**	**AUC**
LogReg	54.91%	0.5277	16.34%	0.4982
MLP	56.13%	0.5689	19.88%	0.5027
DAM-p_u	**79.25%**	0.8763	43.30%	0.6680
DAM-p_u,p_s	79.21%	**0.8797**	**43.80%**	**0.6715**

Comparison of models that utilize only structured data, against DAM models trained on only unstructured, and on both types of input data.

- **Conclusion**:
  - Using **text** data **only** vastly **outperforms** efforts of using only **structured** data on both binary and multi-class classification problem
  - **Combining** the two data sources yields the **best** results overall

Research Question 2: state-of-the art DL models for processing model

- **Q**: How does the DAM model compare to the sota baseline models on binary classification task

(2 resources need as positive class or $\leq$ 2 resources needed as negative class)

	Accuracy		**ROC AUC**	
Models	p_u	(p_u, p_s)	p_u	(p_u, p_s)
embd	55.30%	64.33%	0.5165	0.6155
bi-LSTM	76.59%	77.08%	0.84626	0.8523
VDCNN	76.81%	77.70%	0.8467	0.8609
DSMP	*78.79%*	*78.63%*	*0.8713*	*0.8717*
DAM	**79.25%**	**79.21%**	**0.8763**	**0.8797**

DAM model vs baselines performances for the binary classification task of identifying resource-intensive patients.

- **Conclusion**: The Deep Attention Model outperforms all other state-of-the-art models on the task of identifying resource-intensive patients

Deep Attention Model for Triage of Emergency Department Patients: Research Question 3

- **Q**: How does the DAM model compare to the sota baseline models on multi-class classification task

(classes represent number of resources category - 0,1,2,3,4,5)?

	Accuracy		Average AUC	
Models	p_u	(p_u, p_s)	p_u	(p_u, p_s)
Embd	14.74%	14.74%	0.5000	0.5000
bi-LSTM	39.16%	38.50%	0.6401	0.6358
VDCNN	*39.68%*	*41.33%*	0.6390	*0.6506*
DSMP	39.61%	40.67%	*0.6412*	0.6494
DAM	**43.30%**	**43.80%**	**0.6680**	**0.6715**

DAM model vs baselines performances for the multi-class classification task of predicting number of resource category.

- **Conclusion**: The Deep Attention Model outperforms all other state-of-the-art models on the task of identifying number of resources needed for a patient

Deep Attention Model for Triage of Emergency Department Patients: Research Question 4

- **Q**: How does the DAM model performance compare to nurses' performance in the number of resources category prediction task?

- Nurses estimate first acuity level (FAL) as an estimate of number of resources cat.
- We approximate FAL to resource categories as follows:

Number of Resources Category	First Acuity Level
0	Level 5
1	Level 4
2 or 3	Level 3
4 or 5	Level 2

Number of Resources Category (NRC)	Fisrt Acuity Level (FAL)					
	Level 2	Level 3	Level 4	Level 5	NRC dist.	Recall
4 or 5	12.6%	14.5%	0.9%	0.1%	28.2%	0.4481
2 or 3	5.0%	21.7%	7.3%	0.7%	34.7%	0.6250
1	1.3%	10.7%	7.6%	1.4%	21.1%	0.3615
0	0.7%	8.0%	5.7%	1.7%	16.0%	0.1045
FAL distribution	20%	55%	22%	4%		
Precision	0.6401	0.3953	0.3542	0.4343	Accuracy	**0.4362**

Confusion matrix: **Nurses** performance

- DAM clearly **outperforms** nurses assessment on this task
- 16% improvement in accuracy

Number of Resources Category (NRC) - True	Number of Resources Category Predictions					
	4 or 5~ Level 2	2 or 3: Level 3	1 : Level 4	0 : Level 5	NRC true dist.	Recall
4 or 5	22.2%	5.4%	1.4%	0.4%	29%	0.7549
2 or 3	7.2%	19.3%	6.9%	2.3%	36%	0.5412
1	1.1%	6.0%	9.1%	4.0%	20%	0.4513
0	0.3%	1.7%	3.8%	9.0%	15%	0.6105
NRC predictions dist.	31%	32%	21%	16%		
Precision	0.7211	0.5965	0.4305	0.5733	Accuracy	**0.5961**

Confusion matrix: **DAM** performance

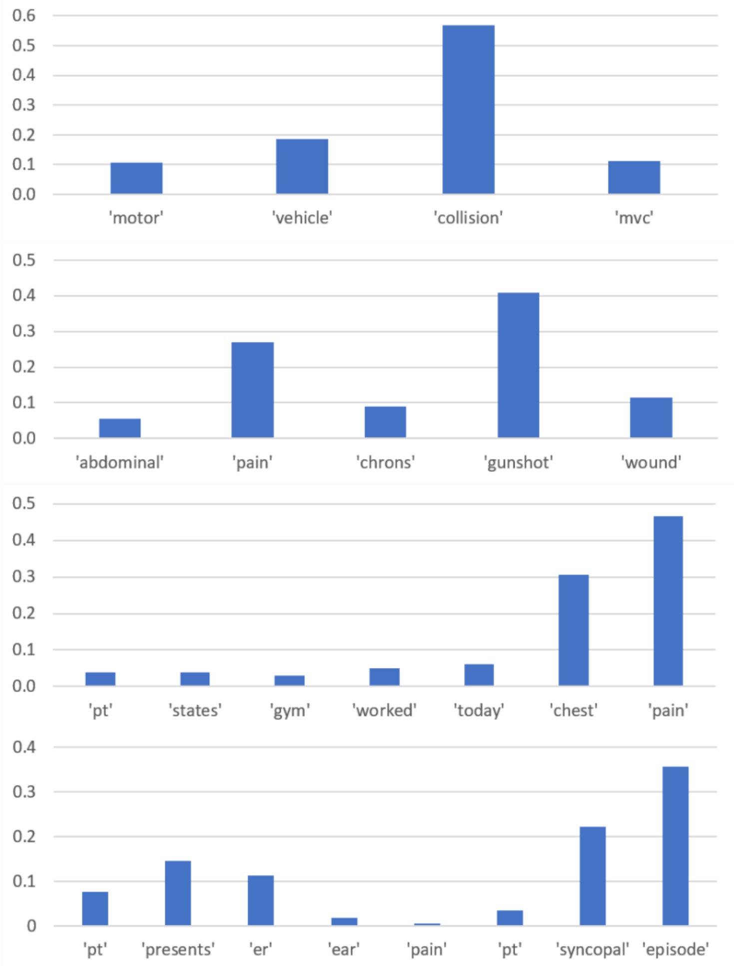
Deep Attention Model for Triage of Emergency Department Patients: Research Question 5 - Interpretability

- **Q**: How can the **attention weights** be utilized by medical practitioners to **understand** predictions made by this model?

- Providing some **intuition** on how was the estimation obtained is **mandatory**

- The DAM's **attentions** learned for words in notes can act as a **proxy** for such intuition.
  - Higher weights can tell triage staff what was factored in the given prediction

- If everything is rendered satisfying by the staff, **actionable decision** can then be taken.



Examples of nurse's notes with attentions

Conclusion

- In this study we addressed the problem of high variance subjective resource utilization outcomes prediction in triage rooms.
- For this task we show that utilizing nurses' notes can provide a significant improvement in accuracy compared to standard continual and nominal data.
- We proposed a novel model to exploit medical texts and obtain state-of-the-art predictive accuracy, finally outperforming reported accuracies of triage staff.
- Attentions the proposed model learns can be very useful in providing clear feedback on what guided the predictions aiding interpretability and clinical acceptance of the model.

Thank you!

Questions?