APCNN: Tackling Class Imbalance in Relation Extraction through Aggregated Piecewise Convolutional Neural Networks

Alisa Smirnova, Julien Audiffren,
Philippe Cudré-Mauroux

eXascale Infolab, University of Fribourg, Switzerland

UNI FR

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

eXascale Infolab

Table of Contents

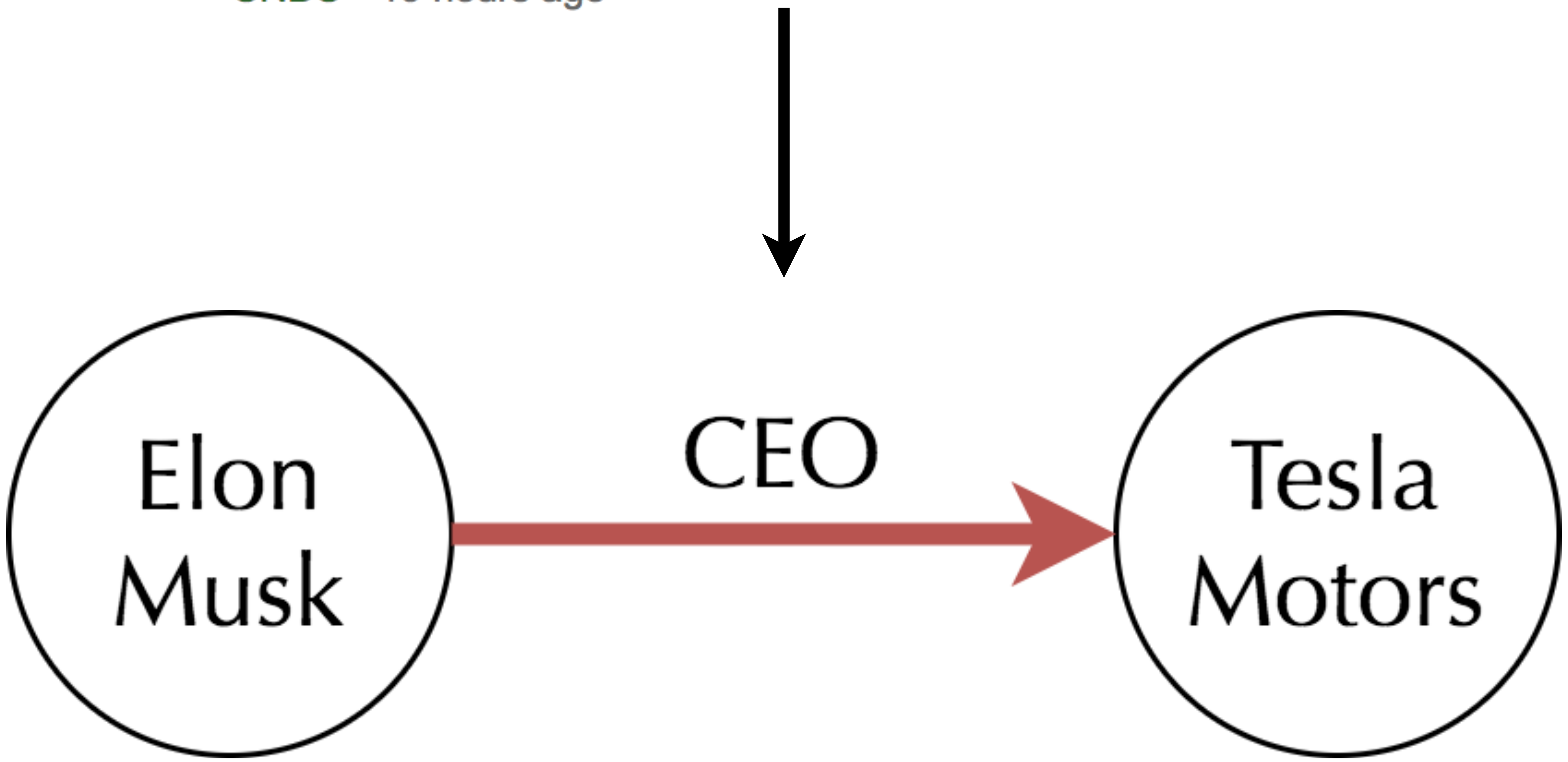
Relation Extraction

Relation extraction is the task of extracting structured information from unstructured text data. Automatically.

Example

Elon Musk bet that Tesla could build the world's biggest battery in ...
Vox - 16 hours ago
Feelings about Tesla CEO **Elon Musk** have always run hot, and lately his company has come in for some withering criticism over (among other things) the working conditions at its factories and its inflated stock valuation. Putting all that aside, though, even his grumpiest critics will have to acknowledge that ...

Analyst jokes **Musk** will make it to Mars with SpaceX before Tesla is ...
CNBC - 19 hours ago



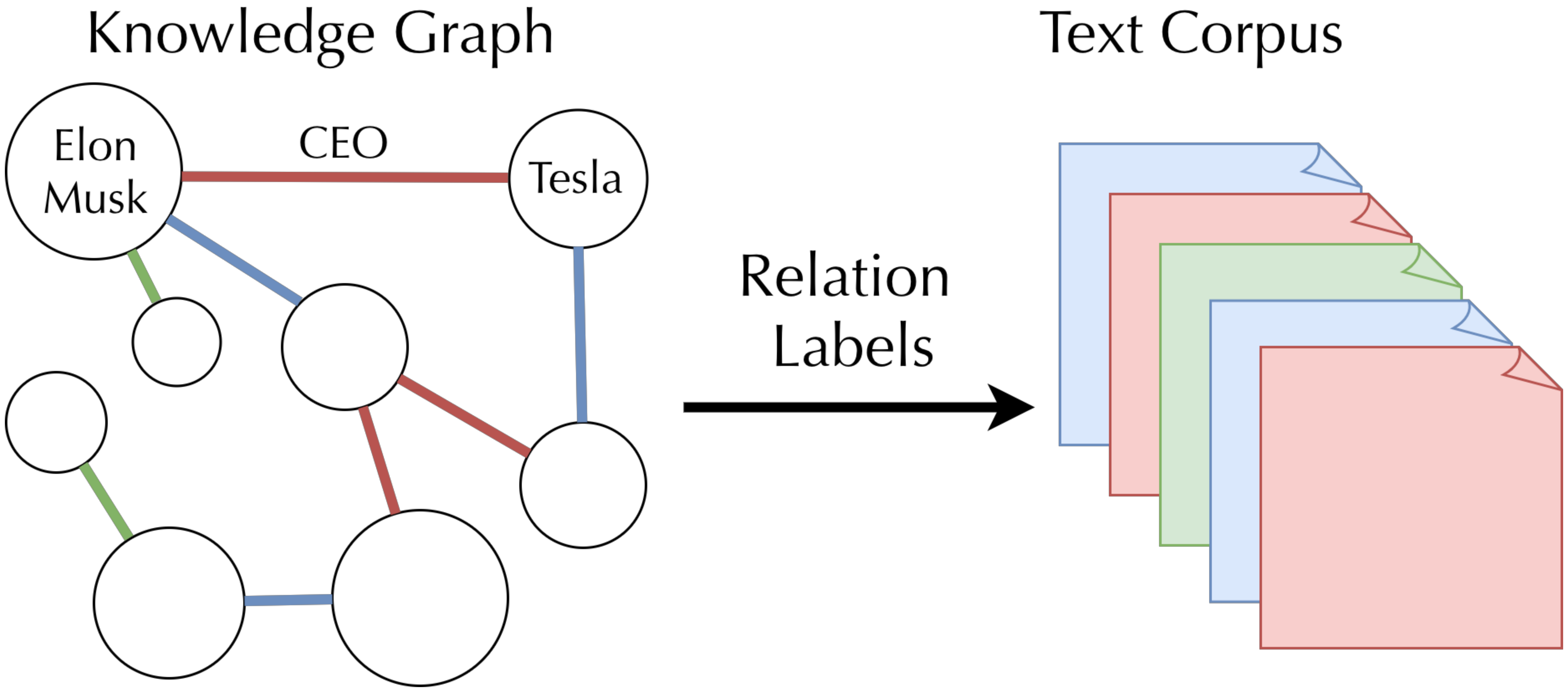
Challenges

- Text corpora nowadays are extremely large.

- Only few annotations are available.

  Distant supervision technique allows to automatically label *any* amount of data.

Distant Supervision



M. Mintz et al. "Distant supervision for relation extraction without labeled data." ACL, 2009.
A. Smirnova and P. Cudré-Mauroux, "Relation extraction using distant supervision: A survey." ACM Computing Surveys, 2019.

Challenges

- Label noise
- Label scarcity
- Label imbalance

Label Noise

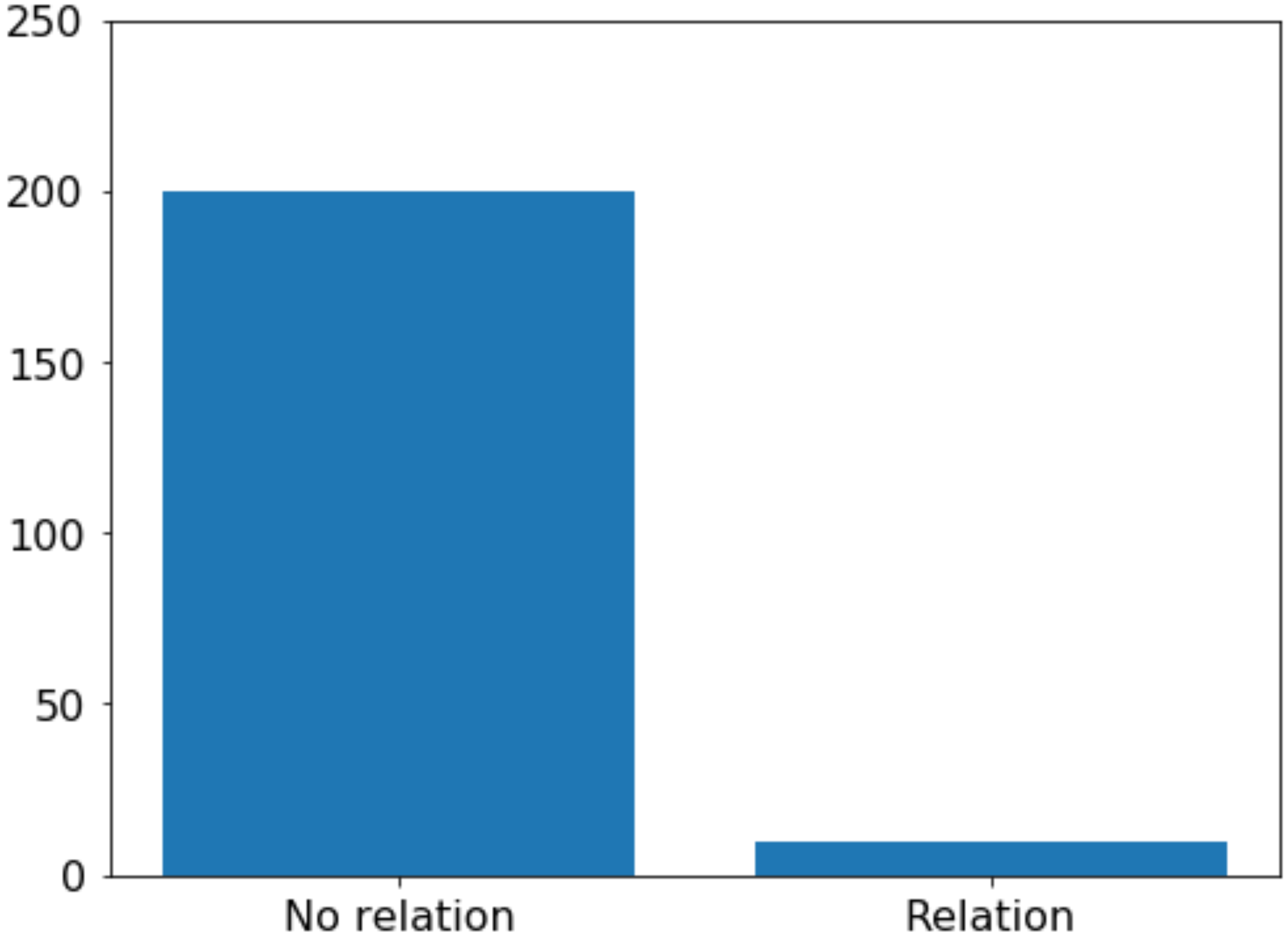
Elon Musk is the co-founder, CEO and Product Architect at **Tesla**.

CEO

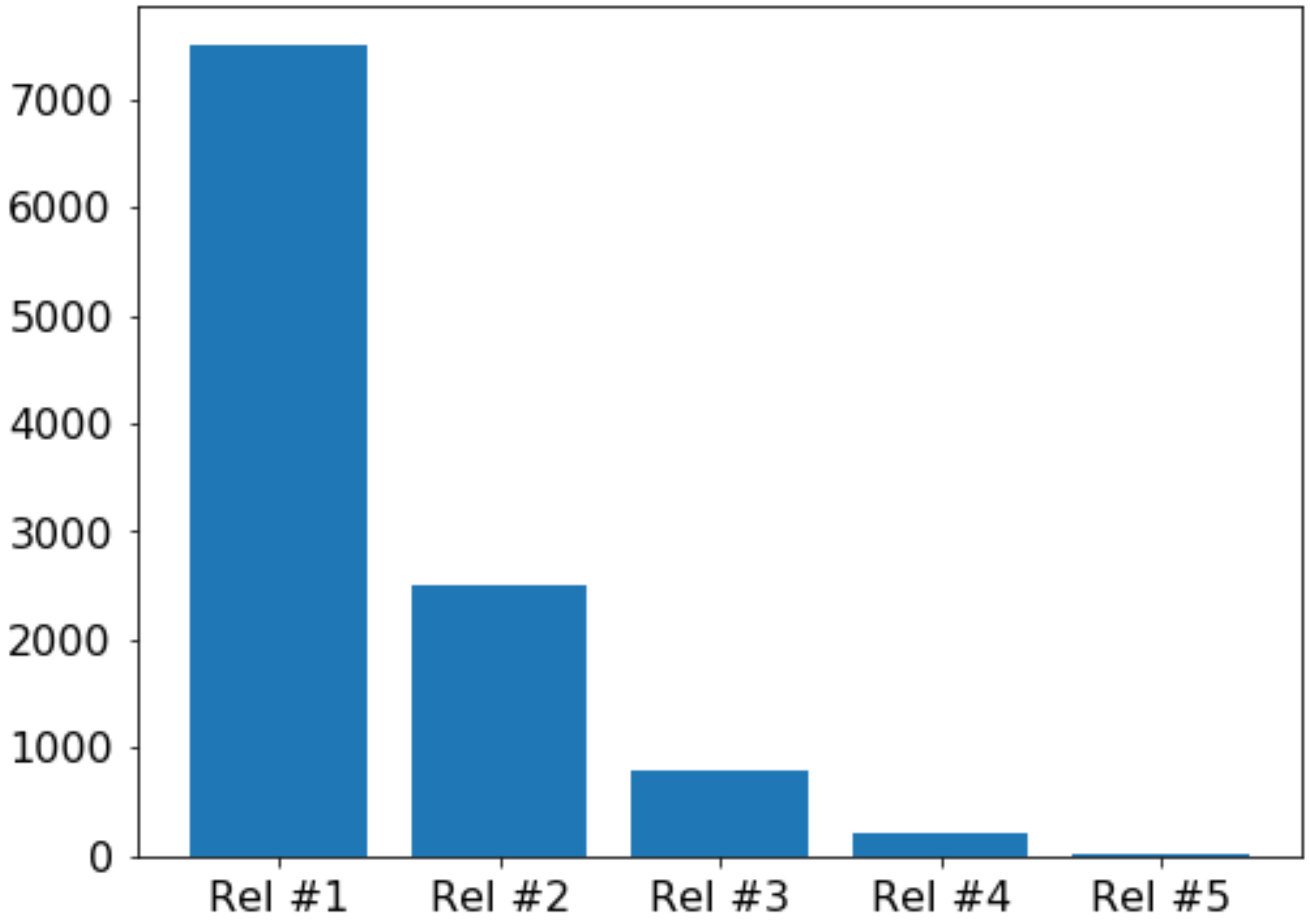
Elon Musk says he is able to work up to 100 hours per week running **Tesla Motors**

?

Label Scarcity



Label Imbalance



Our Approach (APCNN)

- Tackles label scarcity problem
- Tackles label imbalance problem
- Takes into account wrong labels

APCNN

The model consists of two sub-models:

- Binary classifier distinguishes “No relation” from “Some relation”.

- Multiclass classifier predicts exact relation label.

- Both sub-models are convolutional neural networks.

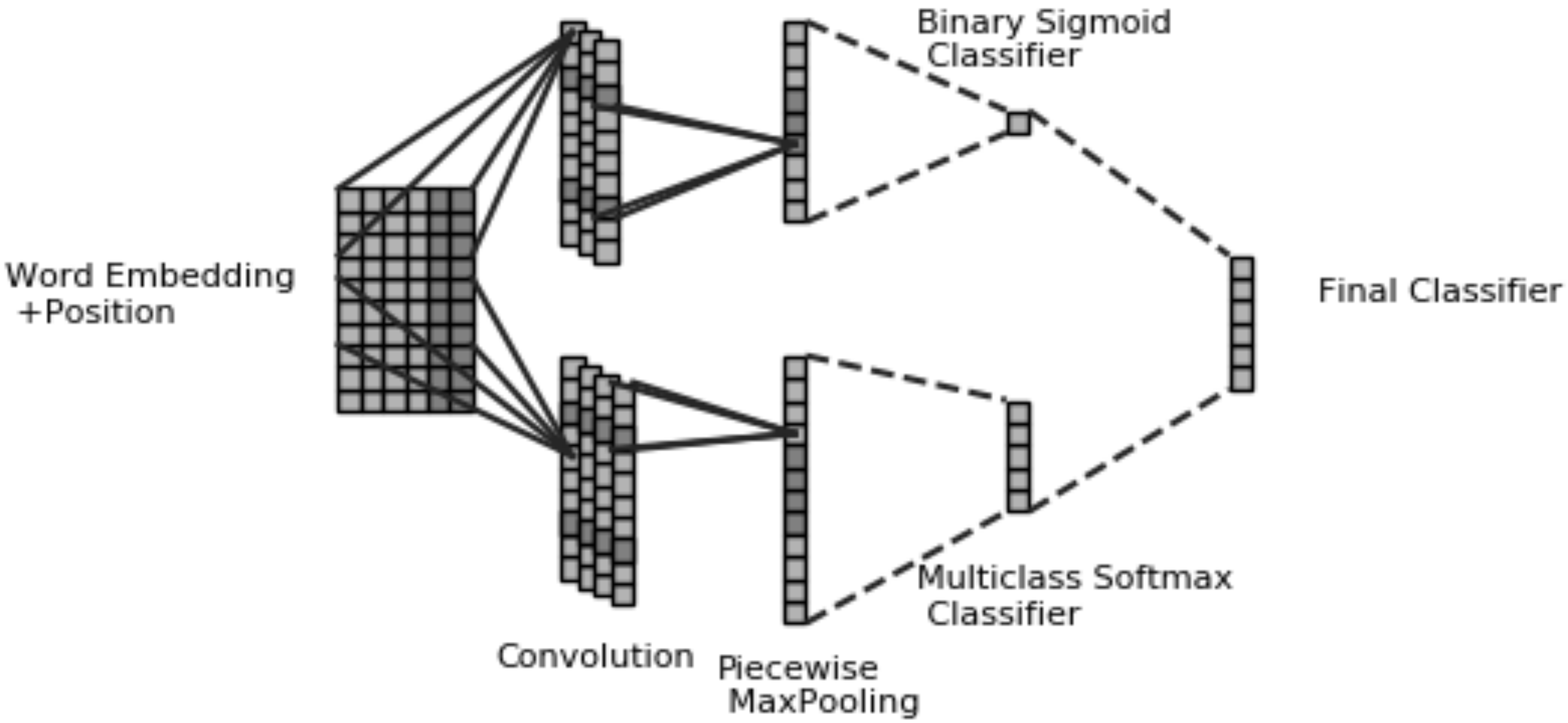
Input of each classifier is a *bag* – a set of all sentences mentioning the same entity pair.

Input Representation

	word embedding					position embedding	
quick						-2	-7
brown						-1	-6
fox						0	-5
jumps						1	-4
over						2	-3
the						3	-2
lazy						4	-1
dog						5	0

For word embeddings we used Word2Vec [T. Mikolov et al., 2013].

Model Architecture



Random Oversampling

- Binary classifier: proportion of positive and negative instances is 1:1.

- Multiclass classifier: proportion of the most frequent relation and the rarest relation is 5:1.

This technique helps tackle both label scarcity and label imbalance.

Loss Function

Ordered weighted average (OWA) of the probabilities of the sentences in bag $\mathscr{B}$ is defined as follows:

$$p_{loss}(r|\mathcal{B}) = \underbrace{(1-\lambda)\max_{s\in\mathcal{B}} p(r|s)}_{\text{sentence maximizing relation}} + \underbrace{\frac{\lambda}{|\mathcal{B}|}\sum_{s\in\mathcal{B}} p(r|s)}_{\text{all sentences}}$$

λ can be interpreted as a weight that we are giving to the sentences in the bag that do not maximize the probability of the relation.

Loss Function

Loss Function for Multiclass classifier is defined as follows:

$$\mathcal{J}(\mathcal{B}) = -\, w_r \log(p_{loss}(r \mid \mathcal{B}))$$

w_r – weight of the relation which is inversely proportional to the size of the class.

Loss Function tackles label imbalance and increases convergence speed.

Predictions

- p_{None} – probability of “None” relation predicted by Binary classifier

- $p(i)$ – probability of relation $i = 1..n$ predicted by Multiclass classifier

Predictions

The final probability distribution $\mathbf{p}(r)$ is defined as follows:

- If $p_{None} > \tau$: $\mathbf{p}_{None} = p_{None}$
- If $p_{None} \leq \tau$: $\mathbf{p}_{None} = \epsilon$
- Probability of relation i: $\mathbf{p}(i) = p_i(1 - \mathbf{p}_{None})$

τ and ϵ are hyperparameters selected by cross-validation.

Evaluation

Two widely used datasets:

- NYTimes (New York Times articles; KG: Freebase)
- Wiki-KBP (Wikipedia articles; KG: Wikipedia Infoboxes)

Metrics used:

- ROC AUC (for binary classification)
- Weighted accuracy and confusion matrix (for overall performance)

Baselines

- PCNN [1]: Piecewise Convolutional Neural Network; uses the same input representation; loss function takes into account only the sentence maximizing the correct relation label.

- CoType [2]: jointly extracts entities and relation using various lexical and syntactic features.

[1] D. Zeng et al. (2015).
[2] X. Ren et al. (2017).

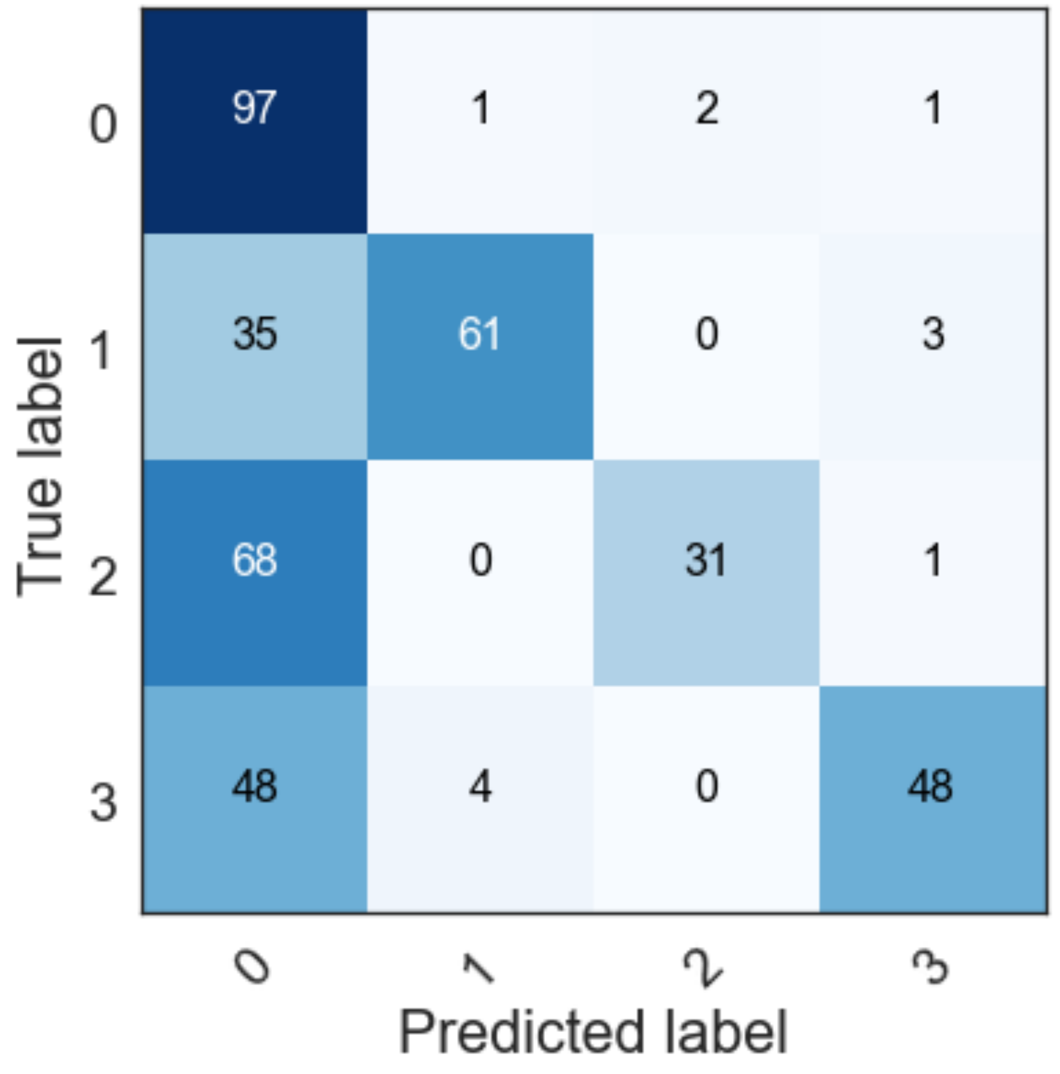
Weighted Accuracy (NYT)

APCNN	25.74%
PCNN	13.47%
CoType	46.03%

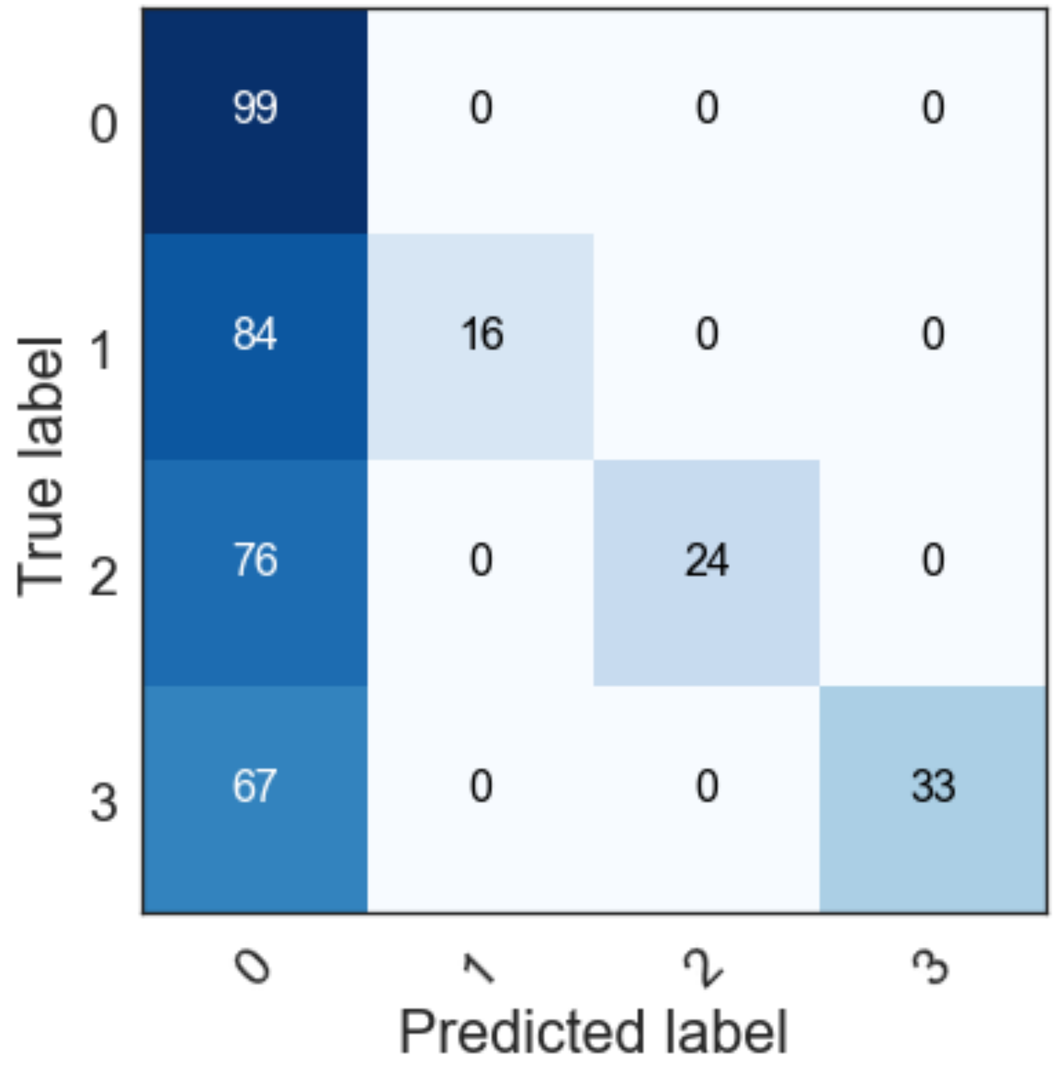
Weighted Accuracy (Wiki)

APCNN	77.70%
PCNN	60.58%
CoType	85.43%

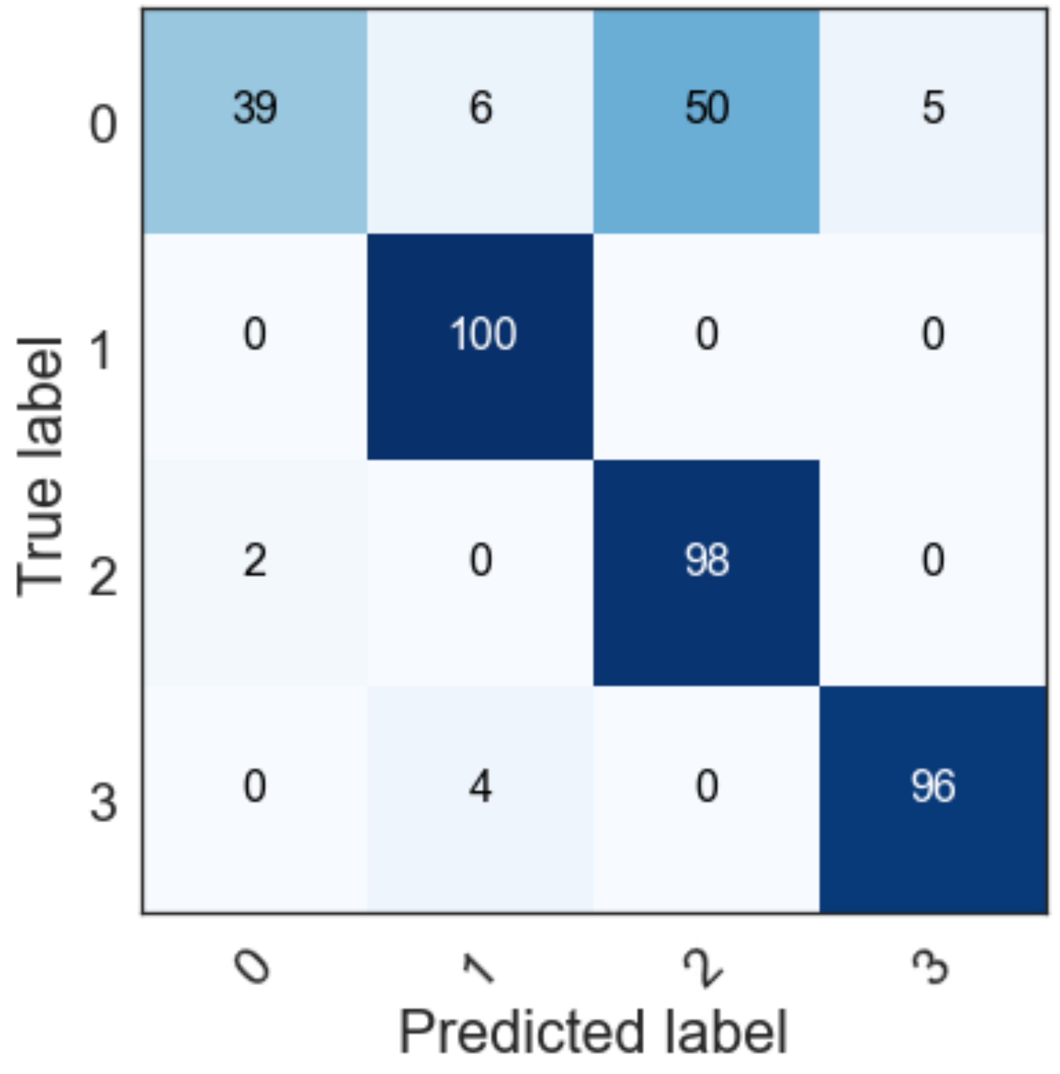
Confusion Matrix



APCNN @ NYT

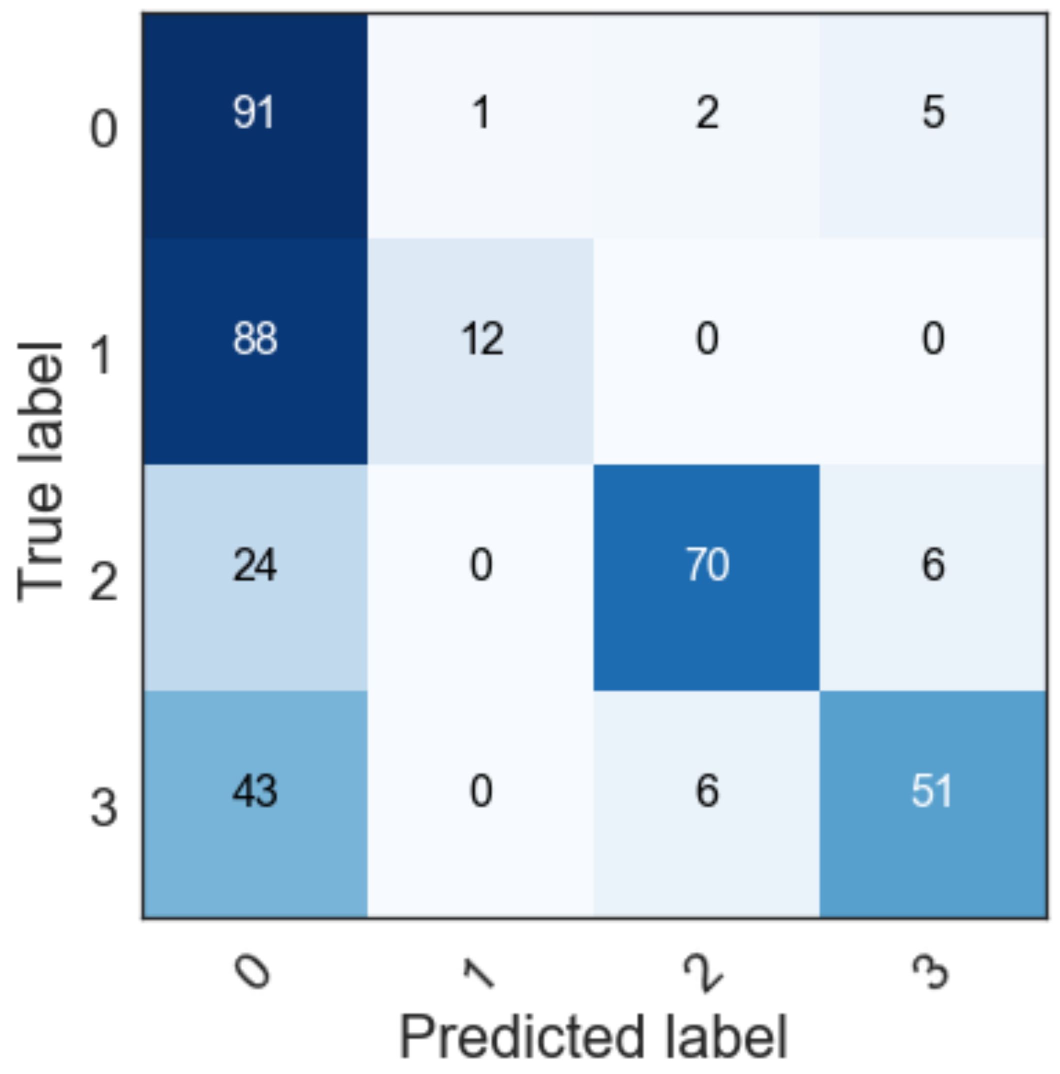


PCNN @ NYT

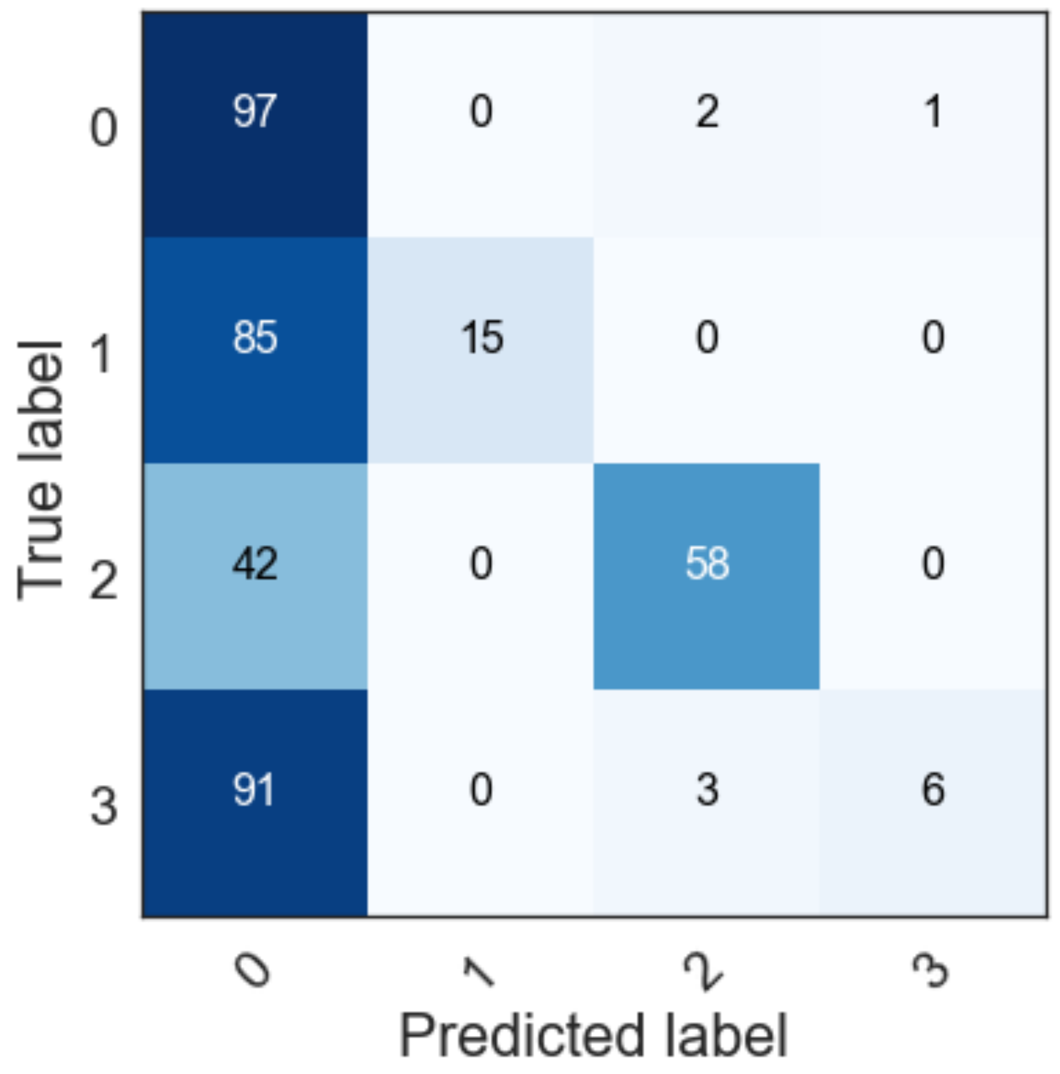


CoType @ NYT

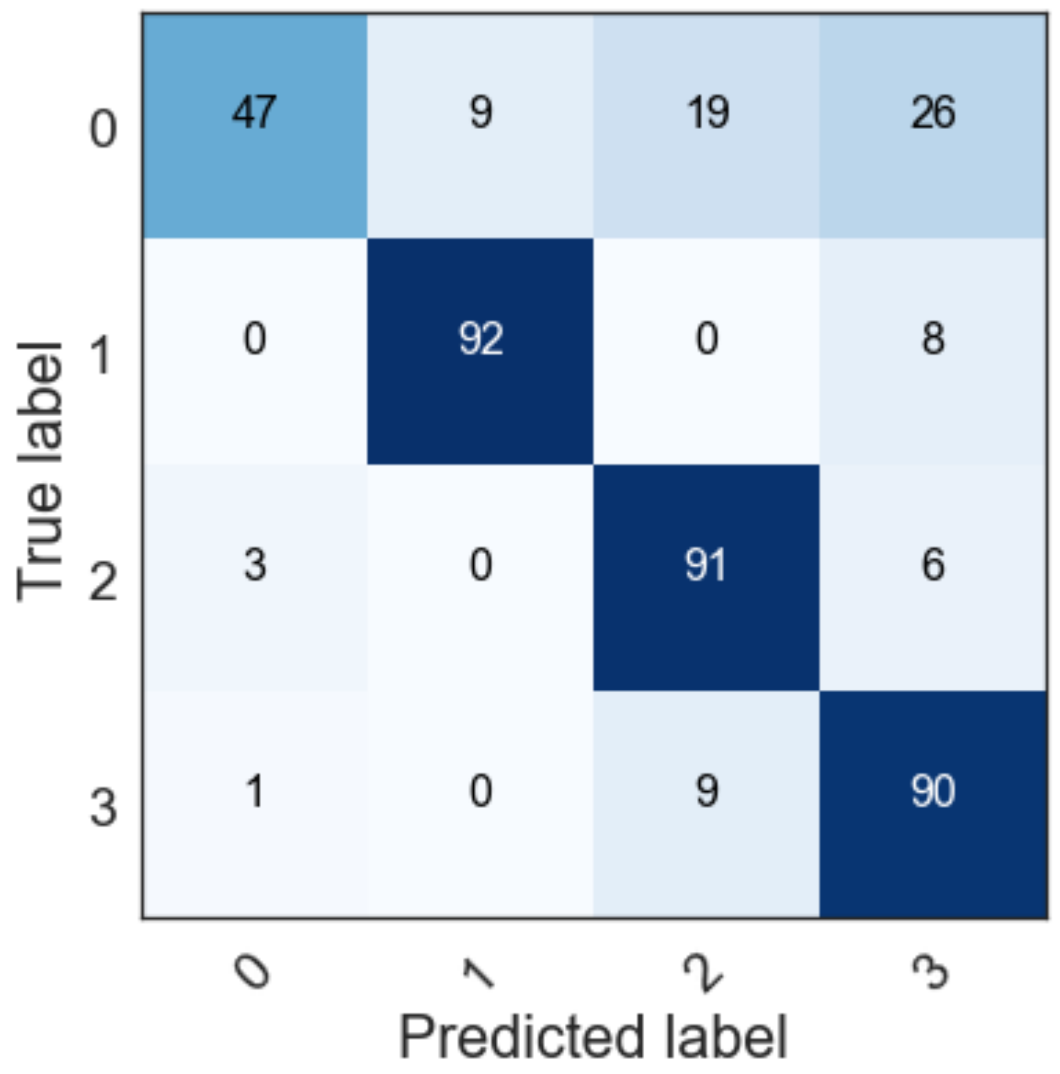
Confusion Matrix



APCNN @ Wiki-KBP



PCNN @ Wiki-KBP



CoType @ Wiki-KBP

Conclusion

- Big challenges in relation extraction are label noise, label scarcity and label imbalance.

- Our model achieves a good balance between predicting the existence of a relation and distinguishing between a set of known relations.

- Future work might include the combination of APCNN and CoType.

Thanks for your attention!