

# CLaC at SMM4H 2020: Birth defect mention detection

Parsa Bagherzadeh, Sabine Bergler

CLaC Labs, Concordia University

Montreal, Canada

{p.bagher, bergler} @cse.concordia.ca

## Abstract

For the detection of personal tweets, where a parent speaks of a child’s birth defect, CLaC combines ELMo word embeddings and gazetteer lists from external resources with a GCNN (for encoding dependencies), in a multi layer, transformer inspired architecture. To address the task, we compile several gazetteer lists from resources such as MeSH and GI. The proposed system obtains .69 for  $\mu$ F1 score in the SMM4H 2020 Task 5 where the competition average is .65.

**Introduction** Tweets potentially offer a record that is distilled from messages that are written for other purposes. They contain many particular, directly observed experiences, which are of great interest to epidemiologists and health monitors. Situating congenital abnormalities (birth defects) in time and space is important to identify causes of birth defects, find opportunities to prevent them, and improve the health of those living with them. Therefore, identifying tweets mentioning birth defect experiences by a family is helpful, which motivated the SMM4H 2020 Task 5 (Klein et al., 2020).

This paper reports our effort toward building a predictive system to address the task. We try to leverage medical annotations, family relation annotations, as well as dependency relations in our model.

**Task description and data** The SMM4H 2020 task 5 is a 3-way classification problem. Class 1 tweets refer to the user’s child and indicate that he/she has a birth defect. Class 2 tweets are ambiguous about whether someone is the user’s child or has a birth defect mentioned in the tweet. Class 3 tweets merely mention birth defects. Training and development sets provided by the organizers include 14705 and 3677 tweets respectively and the evaluation set includes 4603 tweets.

**Preprocessing and annotations** Tweets are tokenized using the ANNIE tweet tokenizer (Cunningham et al., 2002) as well as the hashtag tokenizer (Maynard and Greenwood, 2014). URLs and user mentions are removed. We use the Stanford parser (Klein and Manning, 2003) to encode dependencies and ANNIE *Names* for named entity recognition.

We compiled gazetteer lists for birth defect detection from MeSH (Lipscomb, 2000), namely *BirthDef*, a list of congenital, hereditary, and neonatal diseases and abnormalities compiled from MeSH C16 and *PregComp*, a list of pregnancy complications compiled from MeSH C13.703. Each MeSH sub-tree is traversed depth-first and all entry terms for each heading (both for internal nodes and leaves) are added to the gazetteers. A word list of family terms, *FamilyRel*, that includes terms like *son*, *daughter*, *cousin*, *etc.* is also compiled. Moreover, a list *Acquaintance* which contains terms like *friend*, *colleague*, *neighbor*, *etc.* is extracted from the General Inquirer (Stone et al., 1966).

These four gazetteer lists form the parameter *Annotation* or  $A = \{BirthDef, PregComp, FamilyRel, Acquaintance\}$  and form the four annotation types.

**Deep model** Our system consists of four stacked layers. Each layer  $l \geq 2$  receives token representations  $h_i^{l-1}$  and passes new representations  $h_i^l$  to the next layer. The layers are:

**Layer 1:** annotation embedding and token embedding. Annotation types are embedded using a matrix  $M \in \mathbb{R}^{4 \times 1024}$  (four rows corresponding to four annotation types).  $M$  is initialized randomly

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

and learned as a parameter during training for the main classification task. To embed tokens we use ELMo (Peters et al., 2018). The new representations  $h_i^1$  is obtained by summing annotation embeddings to token embeddings (see Figure 1).

**Layer 2:** the encoder part of the Transformer (Vaswani et al., 2017). The encoder gets the representations  $h_i^1$  and outputs representations  $h_i^2$ . The number of heads in the multi-head attention is  $n_{heads} = 4$  and the dimensionality of the feed-forward layer is  $d_{FF} = 1024$ .

**Layer 3:** graph convolutional network (GCNN) (Kipf and Welling, 2017) for dependency relation encoding following (Marcheggiani and Titov, 2017). In GCCN each token is represented based on its adjacent tokens in dependency parse by  $h_i^l = ReLU(\sum_{j \in \mathcal{N}(i)} W_{L(i,j)} h_j^l + b)$  where  $\mathcal{N}(i)$  is the set of tokens adjacent to token  $i$  and  $L(j, i)$  is the label of the arc from token  $j$  to token  $i$ . Note that the network is not tied, i.e.  $W_{L(i,j)}$  depends on the arc labels. GCNN receives  $h_i^2$  and outputs token-wise representations  $h_i^3$ .

**Layer 4:** attention (Bahdanau et al., 2015), which calculates importance scores  $e_i = w_{att}^T h_i^3$  using a latent context vector  $w_{att}$  and normalizes the scores using softmax ( $\alpha_i = \frac{\exp(e_i)}{\sum_j e_j}$ ) for a weighted sum  $H = \sum_i \alpha_i * h_i^3$ .

Attention produces an output vector  $H \in \mathbb{R}^{1024}$  for the tweet, which is fed into three decision neurons for classifying into the three output classes.

Our model is implemented using PyTorch (Paszke et al., 2017) and optimized with the Adam optimizer (Kingma and Ba, 2015) with a learning rate of  $lr = 0.0005$  for 7 to 10 epochs.

	$h_1^1$ =	$h_2^1$ =	$h_3^1$ =	$h_4^1$ =	$h_5^1$ =	$h_6^1$ =	$h_7^1$ =	$h_8^1$ =	$h_9^1$ =
Annotation Embedding	$M_{Name}$	$\mathbf{0}$	$M_{BirthDef}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
	+	+	+	+	+	+	+	+	+
Token Embedding	$E_{Milo}$	$E_{has}$	$E_{Hydrocephalus}$	$E_{causing}$	$E_{him}$	$E_{to}$	$E_{need}$	$E_{brain}$	$E_{shunt}$
Input	Milo	has	Hydrocephalus	causing	him	to	need	brain	shunt

Figure 1: Additive annotation embedding.  $M_a$  is the row in  $M$  that corresponds to annotation type  $a \in A$

**Submitted system** Due to a submission issue, only one of three planned configurations was submitted.

Ablation studies on our validation data showed that all four layers of the architecture add to the overall performance:

Table 1: Ablations and performance. TE denotes token embedding, AE denotes annotation embeddings

System	Layer	Validation set						Test set		
		Class 1 F1	Class 2 F1	$\mu P$	$\mu R$	$\mu F1$	$\mu P$	$\mu R$	$\mu F1$	
TE+Trans	1,2,4	.70	.63	.72	.60	.65	-	-	-	
TE+GCNN	1,3,4	.72	.64	.74	.60	.67	-	-	-	
TE+AE+Trans	1,2,4	.75	.65	.70	.69	.70	-	-	-	
<b>TE+AE+Trans+GCNN</b>	1,2,3,4	.76	.68	.72	.69	.71	.71	.67	.69	
competition mean	-	-	-	-	-	-	.62	.68	.65	

The results on our development set suggest that the best performance is achieved by the full system (indicated in boldface). The official competition results are provided on the right side of Table 1. The micro average scores for our submission is close to its development score, which we consider an important sign of robustness in our system.

**Conclusion** For the SMM4H 2020 Task 5 competition we proposed a multi-layer system to leverage gazetteer list annotations as well as a dependency parse. The reported experiments here suggest that incorporating external knowledge in form of textual annotation has the potential to enhance the performance of models trained on moderate-sized training sets in a robust and efficient manner.

## References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR'15*.
- H Cunningham, D Maynard, K Bontcheva, and V Tablan. 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. In *ACL'02*.
- DP Kingma and J Ba. 2015. Adam: A method for stochastic optimization. In *3Proceedings of the 3rd International Conference for Learning Representations (ICLR 2015)*. arXiv:1412.6980 [cs.LG].
- Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *In Proceedings of ICLR*.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics*.
- Ari Z. Klein, Ivan Flores, Arjun Magge, Anne-Lyse Minard, Karen O'Connor, Abeed Sarker, Elena Tutubalina, Davy Weissenbacher, and Graciela Gonzalez-Hernandez. 2020. Overview of the fifth Social Media Mining for Health Applications (SMM4H) Shared Tasks at COLING 2020. In *Proceedings of the Fifth Social Media Mining for Health Applications (SMM4H) Workshop & Shared Task*.
- Carolyn E Lipscomb. 2000. Medical subject headings (MeSH). *Bulletin of the Medical Library Association*, 88(3).
- Diego Marcheggiani and Ivan Titov. 2017. Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515.
- DG Maynard and MA Greenwood. 2014. Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. In *LREC 2014*.
- Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. In *NIPS 2017*.
- Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of NAACL-HLT*, pages 2227–2237.
- Philip J Stone, Dexter C Dunphy, and Marshall S Smith. 1966. *The general inquirer: A computer approach to content analysis*. MIT press.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*.