Mapping layperson medical terminology into the Human Phenotype Ontology using Neural Network models

Enrico Manzini[1,2,3], Jon Garrido-Aguirre[1,2,3], Alexandre Perera-Lluna[1,2,3]

[1]B2SLab, Departament d'Enginyeria de Sistemes, Automàtica i Informatica Industrial, Universitat Politècnica de Catalunya, Barcelona, Spain
[2]Networking Biomedical Research Centre in the subject area of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain
[3]Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain

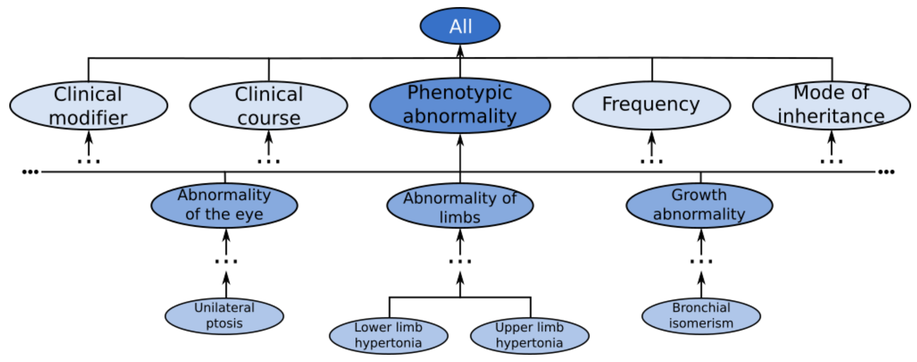
Introduction

The precise analysis of the clinical phenotypes of an individual is known as deep phenotyping, a methodology with the potential to improve the identification of disease with prognostic and therapeutic implications [1]. An essential tool for deep phenotyping is the Human Phenotype Ontology (HPO), a standardized vocabulary of human phenotypic abnormalities [2]. The terminological gap that exists between laypeople (i.e. patients) and HPO, hinders its use for deep phenotyping outside clinical and academic contexts [3], for instance in patient-driven initiatives e.g. Share4Rare. The main goal of this work is to provide an instrument based on neural network algorithms for the automatic translation between layperson terminology and the specific vocabulary of HPO. This way, given a layperson term (e.g. dropping of one upper eyelid) the model returns the corresponding HPO term (unilateral ptosis, `HP:0007687`).

Materials: The Human Phenotype Ontology (HPO)

Train and test sets for the models were created using layperson terms [4] (i.e. synonyms) and descriptions of HPO classes present in HPO (release `2018-12-21`), excluding obsolete terms. These inputs were mapped in $\mathbb{R}^{55}$ with a text tokenizer. Sentences longer than 55 words were excluded to avoid sparsity.

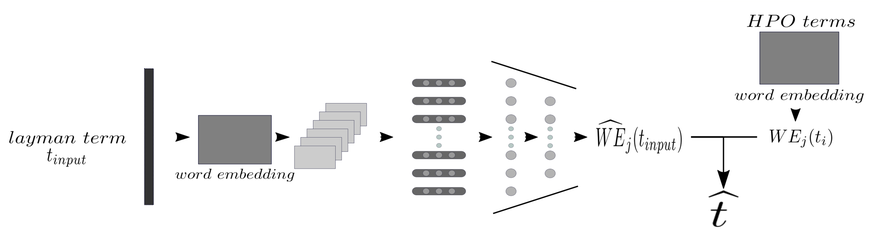
Id: HP::0007687
Name: "Unilateral ptosis"
Description: "A unilateral form of ptosis."
Synonyms: "Dropping of one upper eyelid"
Is a: "Ptosis"
Xref: UMLS:C1866806;



Methods: Models

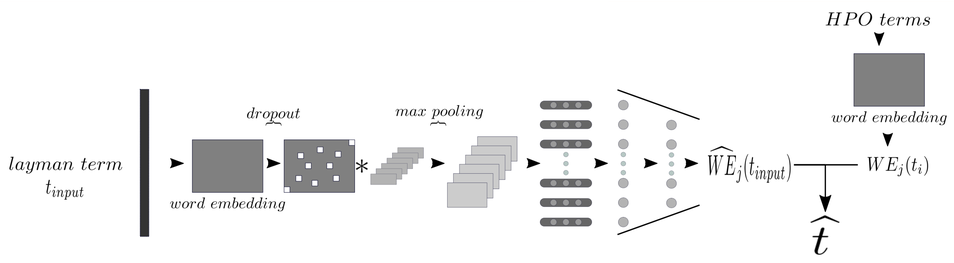
LSTM-D (arch. v2)

Input embedding | LSTM | Dense | Output projection



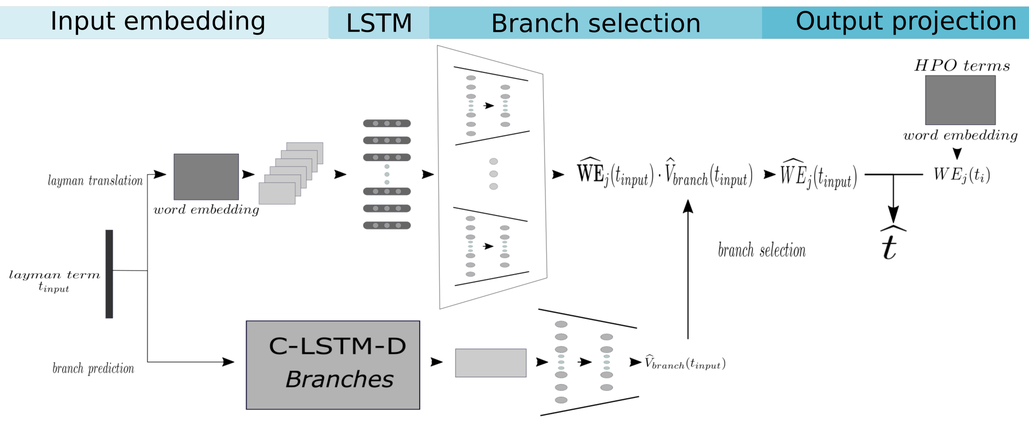
C-LSTM-D (arch. v3)

Input embedding | Convolutional stage | LSTM | Dense | Output projection



LSTM-P (arch. v4)

Input embedding | LSTM | Branch selection | Output projection



Methods: Word Embeddings (WEs)

Generic embeddings. Pre-trained on medical texts from MEDLINE and PubMed [5]. We define the following HPO term representations:

WE_{G1}:

$$WE_{G1}(HPO_i) = \sum_{w_j \in R_t(i)} \vec{w}_j$$

WE_{G2}:

$$WE_{G2}(HPO_i) = \sum_{w_j \in R_t(i)} tf\text{-}idf(w_j) \cdot \vec{w}_j$$

WE_{G3}:

$$WE_{G3} = \sum_{w_j \in R_t^*(i)} f(w_j) \cdot \vec{w}_j$$

with:

$$f(w_j) = \begin{cases} 1 & if \;\; w_j \in R_t(i) \\ e^{-\lambda \cdot j} & if \;\; w_j \notin R_t(i) \end{cases}$$

R_t^* contains the words in R_t together with a list of words related to the HPO term extracted from Wikipedia pages and ordered according the specificity of each word as in [6].

List of words in the i-th HPO term: $R_t(i) = (w_1, w_2, ...w_n)$, $i \in (1, ..., |HPO|)$. Vector representation of word w in the word embedding: $\vec{w}$ New vector representation of the i-th HPO term: $WE_j(HPO_i)$

HPO-specific embedding a.k.a. WE_{LSA}. LSA of a corpus of documents i.e. name, description, parents, and synonyms of each HPO term. Term-document matrix weighted with $tf\text{-}idf$. Dimensionality reduction by SVD.

Combined embedding a.k.a. WE_{SVD}. Meta-embedding proposed in [7]. Concatenation of generic (WE_{G1}) and specific (WE_{LSA}) embeddings. Dimensionality reduction by SVD.

Methods: Training & Evaluation

29,625 terms on the training set, 605 on the test set. 10-fold CV.

$$sim(t_1, t_2) = 1 - \frac{ic(t_1) + ic(t_2) - 2 \times sim_{res}(t_1, t_2)}{2}$$

with:

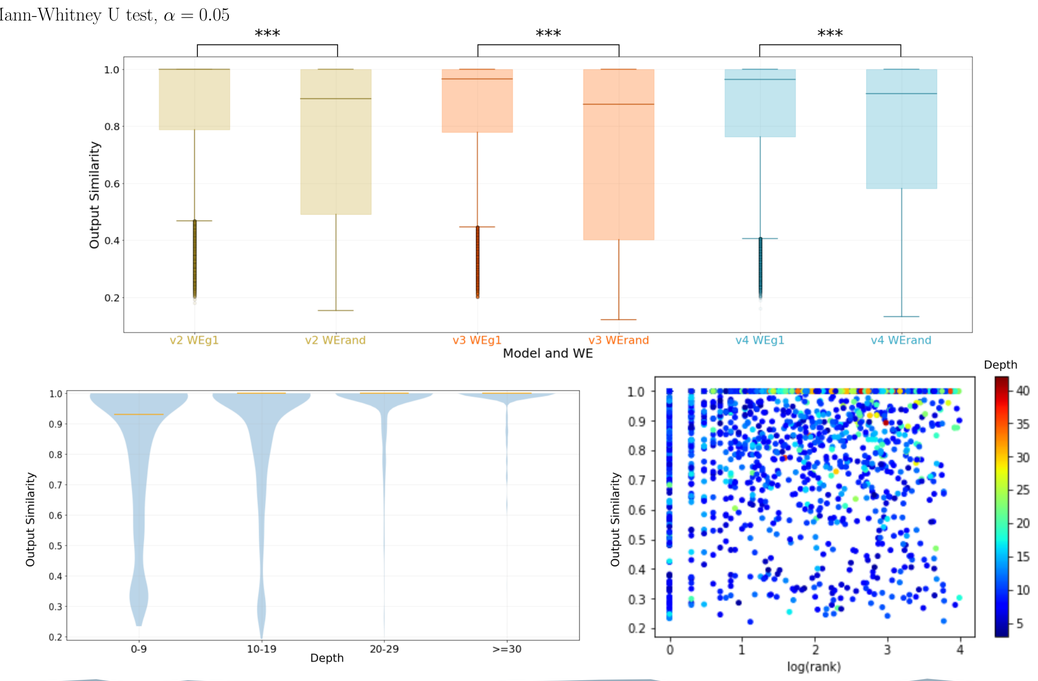
$$ic(t_i) = 1 - \frac{log(depth(t_i) + 1)}{log(|HPO|)} \qquad sim_{res}(t_1, t_2) = \max_{\tau \in S(t_1, t_2)} ic(\tau)$$

Results

	(Intercept)	architectures		LSTM	word embedding				compression		
		v2	v4		g1	g2	g3	LSA	SVD	3:2	2:1	3:1
Estimate	0.19	0.48	0.68	0.19	**2.78**	**1.98**	1.48	**2.48**	**2.96**	0.08	0.11	0.22
y	0.55	0.66	0.70	0.59	**0.95**	**0.90**	0.84	**0.94**	**0.96**	0.57	0.57	0.60
p-value	0.82	0.51	0.37	0.76	**0.03**	**0.04**	0.07	**0.03**	**0.03**	0.92	0.89	0.80

Null model: arch. v3, LSTM 400 units, random WE, no compression; $\alpha = 0.05$

Mann-Whitney U test, $\alpha = 0.05$



Results: Translation Examples

Input	Prediction	Correct	*sim*	% in validation
Absent kidney on one side	Unilateral renal agenesis	Unilateral renal agenesis	1	51.28
Dysharmonic skeletal maturation	Delayed skeletal maturation	Dysharmonic bone age	$0.7 \leq sim < 1$	30.28
Urgency frequency syndrome	Bowel urgency	Urinary urgency	$sim < 0.7$	18.44

Conclusions & Future Work

An explanatory model describing the output similarity in terms of the architecture components and WE types highlights the importance of the WEs in model performance. Whereas the different models show similar performance, the results for `v4` suggest that this architecture contributes the most to the translation of layperson terms, although the combination with certain WEs seems to undermine its influence. In addition, higher input-output compression is beneficial for model performance.

In general, the more specific a term is the better the performance. We consider this to be a promising behaviour: as the more specific a term is the most likely a patient will not know it, it is more important that the models work better with more specific terms. Since direct projections on WEs behave in the opposite way it is presumable that these results are due to a positive contribution of the architectures.

We aim to include the trained translation models in a user-friendly HPO annotator to improve the coverage of the phenotypic profiles of patients in self-declaration of clinical phenotypes.

References

[1] Peter N. Robinson. "Deep phenotyping for precision medicine". In: *Hum Mutat.* 33.5 (2012), pp. 770–780. DOI: 10.1002/humu.22080.
[2] S. Köhler et al. "The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data". In: *Nucleic Acids Research* 42 (2014), pp. D966–974. DOI: 10.1093/nar/gkt1026.
[3] N. Vasilevsky et al. "Plain-language medical vocabulary for precision diagnosis". In: *Nature Genetics* 50 (2018), pp. 474–476. DOI: 10.1038/s41588-018-0096-x.
[4] Nicole A. Vasilevsky et al. "Enhancing the Human Phenotype Ontology for Use by the Layperson". In: *International Conference on Biological Ontology & BioCreative*. 2016. DOI: 10.7490/f1000research.1111752.1.
[5] R. McDonald et al. "Deep Relevance Ranking Using Enhanced Document-Query Interactions". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*, Brussels, Belgium, 2018. arXiv: 1809.01682 [cs.IR].
[6] Mohammad Pilehvar et al. "Improved Semantic Representation for Domain-Specific Entities". In: *Proceedings of the 4th BioNLP Shared Task Workshop*. 2016, pp. 12–16. DOI: 10.18653/v1/W16-2902.
[7] Wenpeng Yin et al. "Learning Word Meta-Embeddings". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin, Germany, 2016, pp. 1351–1360. DOI: 10.18653/v1/P16-1128.