CoDE: Learning Composable Dependency Embeddings

UNIVERSITY OF SUSSEX

Lorenzo Bertolini Julie Weeds David Weir

Text Analytics Group (TAG) – University of Sussex

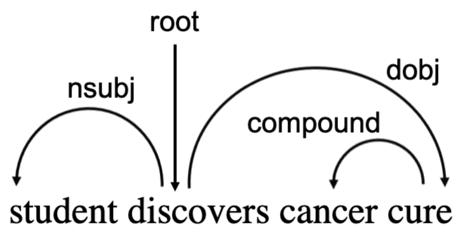
{l.bertolini, j.e.weeds, d.j.weir}@sussex.ac.uk

IWCS 19 VSDD WS

1. Dependency Embeddings

Syntactic representation at word level: what have models encoded?

Previous work mainly focused on expanding word2vec to use first and second order dependencies as contexts, instead of proximity (e.g. [3], [1]).

a) 

b)

Represent. \ Types	Proximity	1^{st} Order Dep.
Untyped	student, cancer, cure	student, cure
Typed	student^{-1}, cancer^{1}, cure^{2}	student$^{\overline{nsubj}}$, cure$^{\overline{dobj}}$

a) Dependency parsed sentence. **b)** Contexts for 'discovers'.

They **generated spaces for word and dependency-aware embeddings**, extensively tested on word-word similarity. **Benefits of and how to use syntax-aware representations for composition remain largely unclear**.

CoDE: from syntax-aware to syntactically-contextualized representations

It has been suggested that syntactic contextualization plays a key role in composition [7]. We introduce a model to generate regular and **syntactically-contextualized word embeddings**, with the clear aim of using the latter **to carry out composition.**

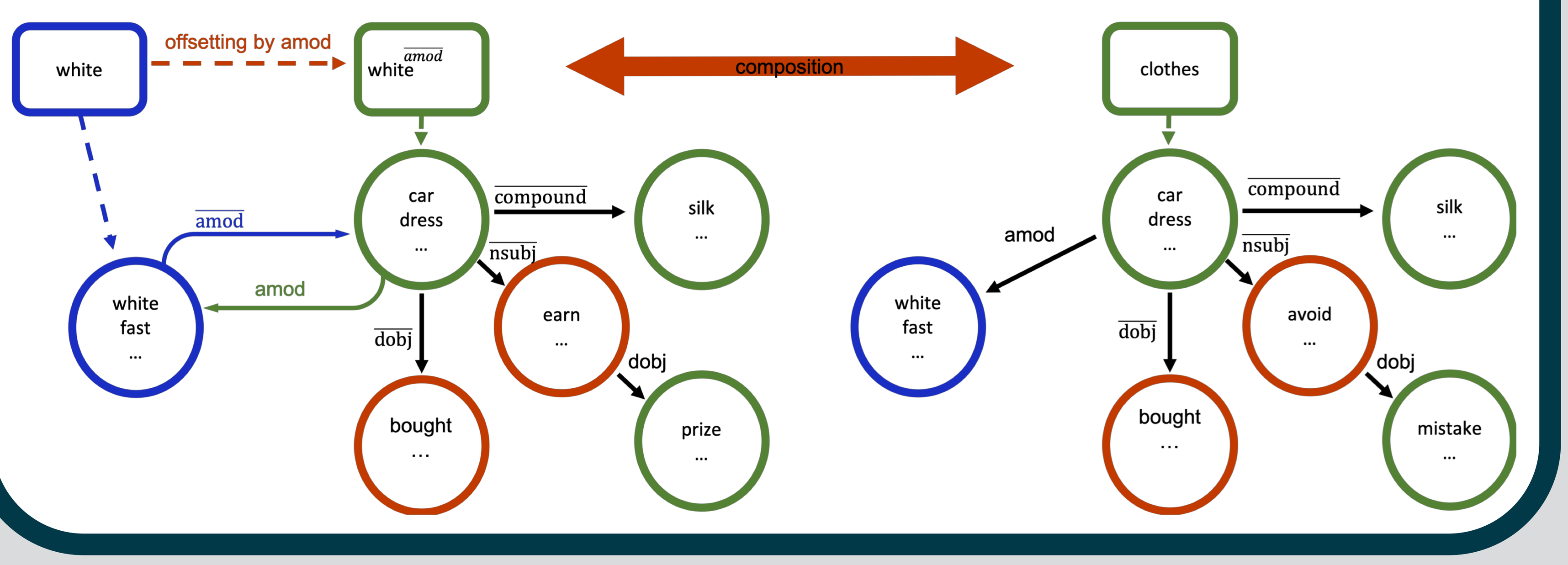
2. APTs

APT: composition via syntactic contextualization in a count-based model

Typed count-based model. Composition is carried out by **contextualising** each **lexeme**, other than the root, **in its syntactic role**, **through** the **offset function**.

Vectors can be seen as collections of dependency trees with the focus (root) on the lexeme they represent. **Offsetting shifts the root along a given path**, allowing feature alignment between words of different types.

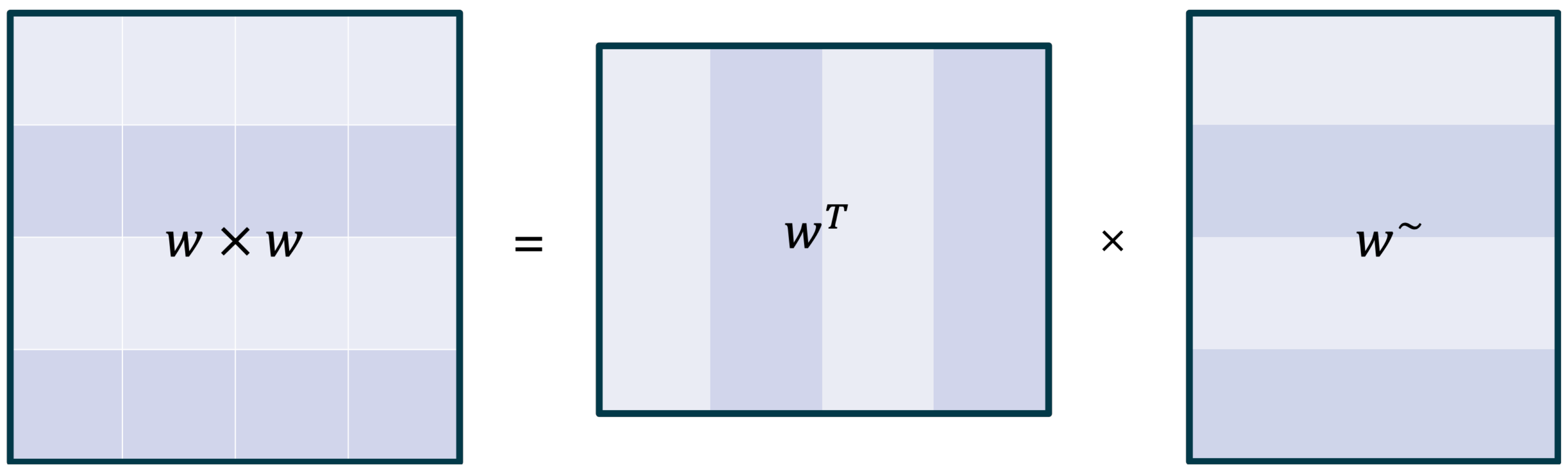
Hence, **offsetting shifts the semantics of the original lexeme**. That is, offsetting *white* by $\overline{amod}$ generates a noun-view of it – i.e. things that are white.



3. CoDE

CoDE: a space for word and composable dependency-embeddings

GloVe [6] optimizes the function $J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - log X_{ij})^2$



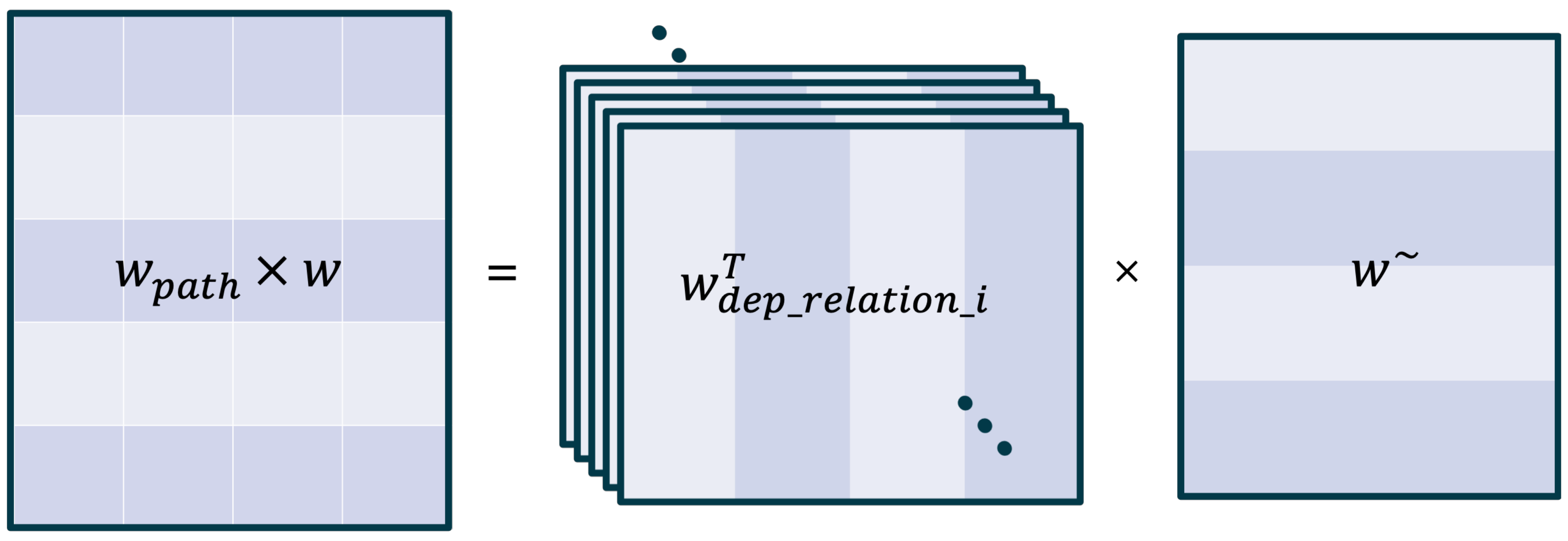
The model learns two vector spaces, defined as focal (w^T) and context ($w^{\sim}$), both of dimensionality vocabulary$\times n$, that are summed after the training.

Our **goal** is to **encode APT's compositionality strategy**, **based on offset representations**, in the embeddings' domain, by learning syntactically-contextualized word representation that will:

1. lie in the same space as regular words
2. exhibit similar features as APT's offset representations
3. hence, improve compositionality

All these aims are linked to the notion of feature alignment in the APT space:

- **to align features, CoDE uses GloVe iteratively**: it creates a different focal sub-space for each dependency relation and trains them by sharing a context one.
- **composition** will be carried out **by adding syntactically-contextualized and regular embeddings**, here seen respectively as offset and root (e.g. white$^{\overline{amod}}$ + clothes, folded + clothes$^{\overline{dobj}}$).



4. Qualitative Investigation

Standard vs. Offset Representations Neighbors Extraction

	white	white $^{\overline{amod}}$	study	study $^{\overline{nmod}}$	write	write $^{\overline{dobj}}$	we	we $^{\overline{nsubj}}$
APT	black red blue yellow green brown dark	shirt stripe pants trousers jacket color coat	research analysis science teach graduate education work	study focus discipline professor institute demonstrate faculty	publish say describe give appear direct produce	write feature include publish have contain make	you I thing one my people this	be opine comment say describe write contend
CoDE context	red black green blue yellow not whitish	pants marking gown tunic hpa lactose blouse	research teach analysis studies science surgeons work	patriarchal asa meta-analysis consular hasidic fraternal 75th	say publish describe co-write give read appear	book novel story script poem song biography	you I they he who it she	do have take go get think make
CoDE full	red black american$^{\overline{amd}.\overline{apps}}$ makeup$^{\overline{nsbj}.det}$ green blue wear $^{dbj.\overline{amd}}$	black $^{\overline{amd}}$ red $^{\overline{amd}}$ solution $^{\overline{nmd}.\overline{cmp}}$ blue$^{\overline{amd}}$ blonde$^{\overline{amd}}$ wear ^{dbj} pants	research correlat.$^{\overline{dbj}.nsbj}$ study^{dbj} demonstrate^{nsbj} reduction $^{\overline{dbj}.nsbj}$ economic $^{\overline{dbj}}$ field^{nmd}	research $^{\overline{nmd}}$ analysis $^{\overline{dbj}}$ scientific$^{\overline{amd}.\overline{nmd}}$ study $^{\overline{nsbj}}$ study$^{\overline{nmd}.\overline{nmd}}$ science^{nmd} clinical$^{\overline{amd}.\overline{nmd}}$	bibliography $^{\overline{dbj}}$ assay $^{\overline{dbj}}$ screenplay $^{\overline{dbj}}$ poetry $^{\overline{dbj}}$ music$^{\overline{dbj}.\overline{cnj}}$ say autobiography $^{\overline{dbj}}$	write^{nsbjpass} publish^{nsbjpass} author^{nmd} publish^{dbj} book write^{nmd} read^{dbj}	you I they lot $^{\overline{dbj}.nsbj}$ thing $^{\overline{dbj}.nsbj}$ get^{nsbj} what $^{\overline{dbj}.nsbj}$	I$^{\overline{nsbj}}$ you$^{\overline{nsbj}}$ they$^{\overline{nsbj}}$ we$^{\overline{nsbj}.xcmp}$ one$^{\overline{nsbj}}$ we $^{\overline{nsbj}.advcl}$ I $^{\overline{nsbj}.ccmp}$

We trained an APT space, and accordingly CoDE, on a clean Wikipedia 2014 corpus, sequentially reduced to ~28k lexemes. Code and embeddings available at github.com/lorenzoscottb/CoDE_iwcs_19 .

5. Composition Task

	APT		CoDE		GloVe
	regular + regular	offset + regular	regular + regular	offset + regular	
Adjective Nouns	.38	.38	.36	**.44**	.38
Compound Nouns	.38	.38	.37	.39	**.47**
Verb Objects	.35	.33	.36	.32	**.39**
Average	.37	.36	.36	.38	**.41**

Mitchell-Lapata 2010 task evaluates similarity between two phrases (e.g. *black hair – dark eyes*). Values report Spearman's ρ. The benchmarks is a pretrained GloVe space (Wikipedia 2014 + Gigaword 5). Both CoDE and GloVe representations have dimensionality of 300.

6. Conclusions

- Novel model to learn standard and dependency-based embeddings that are syntactically-contextualized and composable.
- Results suggests we successfully encoded desired features in representations.
- Further work will focus on composition above the phrase level.

References

1. Komninos, A. & Manandhar, S. (2016). Dependency Based Embeddings for Sentence Classification Tasks. *In North American Chapter of the Association for Computational Linguistics*, 1490–1500.
2. Lapata, M., & Padó, S. (2007). Dependency-Based Construction of Semantic Space Models. *Computational Linguistics*, *33*, 161–199.
3. Levy, O., and Goldberg, Y. 2014. Dependency-based word embeddings. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*
4. Li, C., Li, J., Song, Y., & Lin, Z. (2018). Training and Evaluating Improved Dependency-Based Word Embeddings. *Aaai*, 5836–5843.
5. Mitchell, J., & Lapata, M. (2010). Composition in Distributional Models of Semantics. *Cognitive Science*, *34*(8), 1388–1429.
6. Pennington, J., Socher R., and Manning C., (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing*, pages 1532–1543.
7. Weir, D., Weeds, J., Reffin, J., & Kober, T. (2016). Aligning packed dependency trees: A theory of composition for distributional semantics. *Computational Linguistics*, *42*(4), 727–761.
8. Yin, Y., Wei, F., Dong, L., Xu, K., Zhang, M., & Zhou, M. (2016). Unsupervised word and dependency path embeddings for aspect term extraction. *International Joint Conference on Artificial Intelligence*, 2979–2985.