Transferring Learned Models of Morphological Analogy

ATA@ICCBR22

Esteban Marquer[1], Pierre-Alexandre Murena[2], Miguel Couceiro[1]

[1]Université de Lorraine, CNRS, LORIA, F-54000, France
[2]HIIT, Aalto University, Helsinki, Finland

Sept. 2022

Outline

Introduction

Analogy detection in morphology

Analogical proportion (AP):
$A : B :: C : D$, read "A is to B as C is to D"

Analogy detection:
Given A, B, C, and D, is $A : B :: C : D$ a valid analogy?

Morphological transformation:
change in the form of word, following morphological rules

Morphology in practice:
suffixes, prefixes, infixes, ...

English
dog : *dog**s*** :: *cat* : *cat**s***
singular → plural, add suffix *-s*

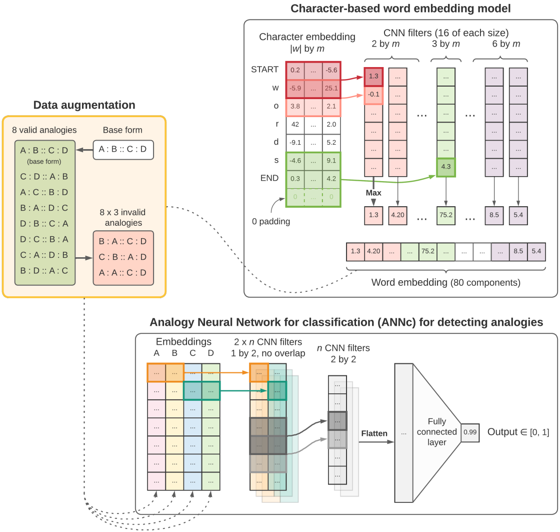
pull : *pull**ed*** :: *achieve* : *achieve**d***
infinitive → past participle, add suffix *-ed*

Motivation

Morphological Analogy:

- benchmark to develop and study models of analogy:
  - textual data: sequential data;
  - interpretable by humans;
  - many languages: similar domains;
  - etc.

Learned models of analogy



ANNc[a]

Analogy Neural Network for classification

In: 4 words A,B,C,D
Out: valid (+)/ invalid (−)

Relies on data augmentation for permutations P^+ and P^-

Relies on axioms for analogy

[a] S. Alsaidi, A. Decker, P. Lay, E. Marquer, P.-A. Murena, and M. Couceiro (2021). "A Neural Approach for Detecting Morphological Analogies". In: *IEEE 8th DSAA*, pp. 1–10

Transferring learned models of analogy

We have a model $M_{\mathcal{D}_{\text{source}}}$ trained on data sampled from a domain $\mathcal{D}_{\text{source}}$
We want to use it on a different domain $\mathcal{D}_{\text{target}}$

- Reduce or eliminate training on $\mathcal{D}_{\text{target}}$
- Improve performance on $\mathcal{D}_{\text{target}}$ using general-purpose model
- Study $M_{\mathcal{D}_{\text{source}}}$, ex: training VS test sets $\leftarrow$ **Ours**

Use $M_{\mathcal{D}_{\text{source}}}$ without change, performance on $\mathcal{D}_{\text{target}}$ depends on (mainly):

- performance on $\mathcal{D}_{\text{source}}$
- how similar $\mathcal{D}_{\text{target}}$ is to $\mathcal{D}_{\text{source}}$ & how well $M_{\mathcal{D}_{\text{source}}}$ generalizes

Motivation

Transfer to:

- study how **data augmentation** impacts ANNc;
- check if ANNc generalizes to a **different dataset**;
- study how ANNc behaves when applied on **different languages**.

Dataset

Siganalogies[1] dataset:

- more than 80 languages in total;
- more than $50\,000$ unique analogies for the ones we use;
- Sigmorphon2016[2] subset: all 10 lang.;
- Sigmorphon2019[3] subset: 42 lang.;
- Japanese Bigger Analogy Test Set[4] subset: Japanese.

---

[1] E. Marquer, M. Couceiro, S. Alsaidi, and A. Decker (2022). *Siganalogies - morphological analogies from Sigmorphon 2016 and 2019*. Version V1

[2] R. Cotterell, C. Kirov, J. Sylak-Glassman, D. Yarowsky, J. Eisner, and M. Hulden (2016). "The SIGMORPHON 2016 Shared Task–Morphological Reinflection". In: *SIGMORPHON 2016*. ACL, pp. 10–22

[3] A. D. McCarthy et al. (2019). "The SIGMORPHON 2019 Shared Task: Morphological Analysis in Context and Cross-Lingual Transfer for Inflection". In: *16th CRPPM workshops*. ACL, pp. 229–244

[4] M. Karpinska, B. Li, A. Rogers, and A. Drozd (2018). "Subcharacter Information in Japanese embeddings: when is it worth it?" In: *RLSNA4NLP*. ACL, pp. 28–37

Data in Siganalogies

Word pairs linked by inflexional morphology transformations:

`revérifier V;SBJV;PST;1;SG revérifiasse`

`tormenter V;SBJV;PST;1;SG tormentasse`

$\rightarrow$ *revérifier* : *revérifiasse* :: *tormenter* : *tormentasse*

Different axioms

Axioms for APs

In previous work[5], we considered the following axioms[6]:

- symmetry (if $A : B :: C : D$, then $C : D :: A : B$);
- central permutation (if $A : B :: C : D$, then $A : C :: B : D$);
- strong reflexivity (if $A : A :: C : D$, then $C = D$);
- strong inner reflexivity (if $A : B :: A : D$, then $B = D$).

Applying them multiple times gives us 8 equivalent forms. We can also derive "forbidden" forms (like $A : A :: C : D$ with $C \neq D$).

[5] S. Alsaidi, A. Decker, P. Lay, E. Marquer, P.-A. Murena, and M. Couceiro (2021). "A Neural Approach for Detecting Morphological Analogies". In: *IEEE 8th DSAA*, pp. 1–10

[6] Y. Lepage (2003). "De l'analogie rendant compte de la commutation en linguistique". FR. Habilitation à diriger des recherches. Université Joseph-Fourier - Grenoble I

3 Axiomatic settings, discussing *central permutation* (CP)

CP subject to discussion.

Example[7]: different domains in the same analogy.

$$2 : 4 :: ab : abab$$

Replace CP by *inside pair reversing* (if $A : B :: C : D$, then $B : A :: D : C$).

[7] Christian Antić (2022). "Analogical Proportions". In: *Annals of Mathematics and Artificial Intelligence* 90.6, pp. 595–644. ISSN: 1573-7470. DOI: 10.1007/s10472-022-09798-y

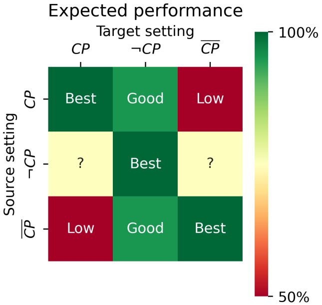
3 Axiomatic settings

Setting	CP	$\neg CP$	$\overline{CP}$
CP results in ... analogies	valid	omitted	invalid
Valid P^+	A:B::C:D C:D::A:B B:A::D:C D:C::B:A **A:C::B:D** **C:A::D:B** **B:D::A:C** **D:B::C:A**	A:B::C:D C:D::A:B B:A::D:C D:C::B:A	A:B::C:D C:D::A:B B:A::D:C D:C::B:A
Invalid P^- A', B', C', D' in P^+	A':A'::C':D' B':A'::C':D' **C':B'::A':D'**	A':A'::C':D' B':A'::C':D'	A':A'::C':D' B':A'::C':D' **C':B'::A':D'** **A':C'::B':D'**

Experiment

Hypothesis: if the model learns the axioms, the performance can be predicted using deductions on the axioms.

Train on a setting and test on another one:



✔: $P^{+}{}_{\neg CP} \subseteq P^{+}{}_{CP}$ and $P^{-}{}_{\neg CP} \subseteq P^{-}{}_{CP}$
$P^{+}{}_{\neg CP} \subseteq P^{+}{}_{\overline{CP}}$ and $P^{-}{}_{\neg CP} \subseteq P^{-}{}_{\overline{CP}}$

✖: $P^{-}{}_{\overline{CP}} \cap P^{+}{}_{CP} \neq \emptyset$

?: $\neg CP$ on unseen permutations.

Balanced accuracy: $\dfrac{A_{P^+} + A_{P^-}}{2}$

Experimental results

Conclusions

Observed results support the hypothesis:
"Z" ✔ goes in the direction that model "learns" axioms based on augmentation

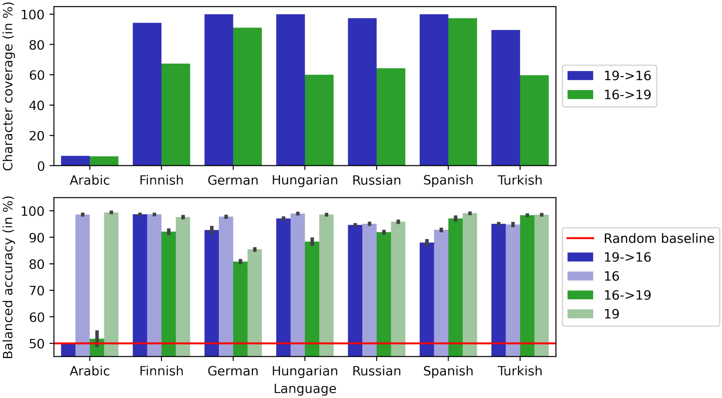
Limitation:
$\overline{CP} \rightarrow CP < CP \rightarrow \overline{CP}$
reason unknown

Different datasets

Different datasets

Languages shared in Sigmorphon2016 and Sigmorphon2019.
Transformations involved can be different, language is the same.

Character coverage: $\frac{|\mathcal{C}_{source} \cap \mathcal{C}_{targets}|}{|\mathcal{C}_{targets}|}$



Conclusion

Model appears to generalize as long as it knows the characters.

Different languages

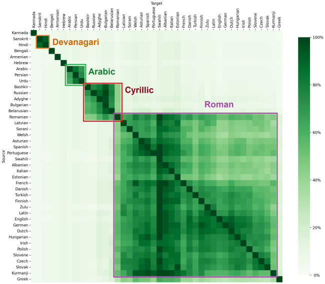
Different languages

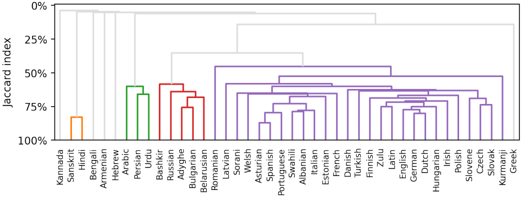
42 “high ressource” languages in Sigmorphon2019.

Character coverage prerequisite for non-zero performance.

Cluster using Jaccard index: $\frac{|\mathcal{C}_{source} \cap \mathcal{C}_{targets}|}{|\mathcal{C}_{source} \cup \mathcal{C}_{targets}|}$

Coverage not a major factor anymore: Pearson correlation of performance coverage $r = 0.0379$ for Roman cluster.



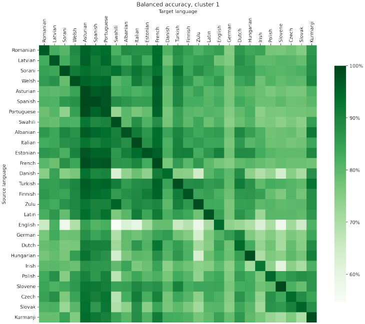


27 languages in the Roman cluster

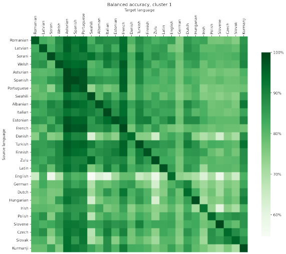
Experimental results

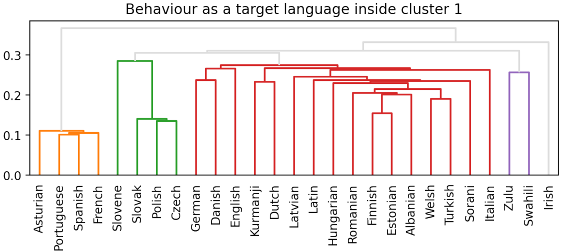
Patterns in performance: **model performance** or **similarities between languages**?

Hypothesis: similarities between languages morphology (vertical bars in matrix).



Experimental results

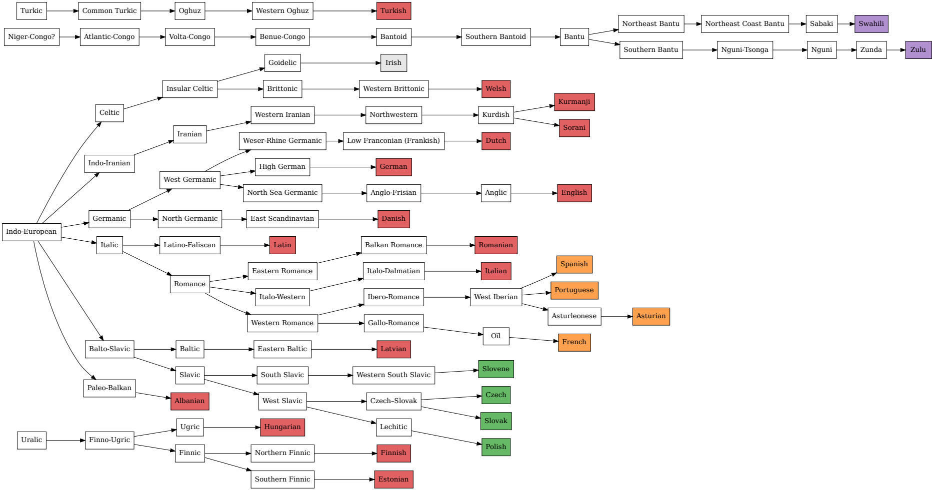




Hypothesis: similarities between languages morphology.

Compare with "language families"
("hierarchical clustering" of languages, "genealogy").

Experimental results



Conclusions

Model transfers well between closely related languages:

- language families to decide of a source language;
- analogy to study morphology.

Questions

Thank you for your attention!

Wikipedia language families

Wikipedia language families

English	
Pronunciation	/ˈɪŋɡlɪʃ/[1]
Ethnicity	English people Anglo-Saxons (historically)
Native speakers	360–400 million (2006)[2] L2 speakers: 750 million; as a foreign language: 600–700 million[2]
Language family	Indo-European • Germanic • West Germanic • North Sea Germanic • Anglo-Frisian • Anglic • **English**
Early forms	Old English • Middle English • Early Modern English
Writing system	Latin (English alphabet) Anglo Saxon runes (historically) English Braille, Unified English Braille
Signed forms	Manually coded English (multiple systems)

Bibliography I

Alsaidi, S. et al. (2021). "A Neural Approach for Detecting Morphological Analogies". In: *IEEE 8th DSAA*, pp. 1–10.

Antić, Christian (2022). "Analogical Proportions". In: *Annals of Mathematics and Artificial Intelligence* 90.6, pp. 595–644. ISSN: 1573-7470. DOI: 10.1007/s10472-022-09798-y.

Cotterell, R. et al. (2016). "The SIGMORPHON 2016 Shared Task–Morphological Reinflection". In: *SIGMORPHON 2016*. ACL, pp. 10–22.

Karpinska, M. et al. (2018). "Subcharacter Information in Japanese embeddings: when is it worth it?" In: *RLSNA4NLP*. ACL, pp. 28–37.

Lepage, Y. (2003). "De l'analogie rendant compte de la commutation en linguistique". FR. Habilitation à diriger des recherches. Université Joseph-Fourier - Grenoble I.

Bibliography II

Marquer, E. et al. (2022). *Siganalogies - morphological analogies from Sigmorphon 2016 and 2019*. Version V1.

McCarthy, A. D. et al. (2019). "The SIGMORPHON 2019 Shared Task: Morphological Analysis in Context and Cross-Lingual Transfer for Inflection". In: *16th CRPPM workshops*. ACL, pp. 229–244.