Insights into Analogy Completion from the Biomedical Domain

Denis Newman-Griffis, Albert Lai, Eric Fosler-Lussier

The Ohio State University National Institutes of Health, Clinical Center
Washington University in St. Louis

BioNLP 2017
August 4, 2017

The Ohio State University
National Institutes of Health Clinical Center
Washington University in St. Louis

TL;DR

Increasing work on training embedding-based models for biomedical applications, but not many resources to evaluate on.

TL;DR

Increasing work on training embedding-based models for biomedical applications, but not many resources to evaluate on.

Analogies have been highly useful in the general domain, so we built an analogy dataset for BioNLP. [1]

[1] https://github.com/OSU-slatelab/BMASS

TL;DR

Increasing work on training embedding-based models for biomedical applications, but not many resources to evaluate on.

Analogies have been highly useful in the general domain, so we built an analogy dataset for BioNLP. [1]

Findings:

- Current embeddings are good at direct chemical/biological relationships, not so good at clinical semantics.
- Changes need to be made to the standard analogy methods to reflect the complexity of real data.

[1] https://github.com/OSU-slatelab/BMASS

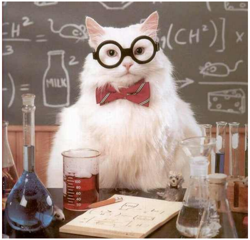
Outline

The analogy completion task

BMASS

How can we make analogies more realistic?

Findings and challenges on our dataset



The analogy completion task

London : England :: Paris : France

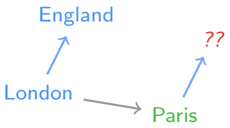
The analogy completion task

London : England :: Paris : ________

The analogy completion task

London : England :: Paris : ________

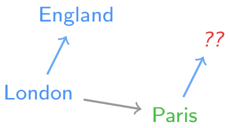
$$d^* = \text{argmax}_{d \in V}\left(\cos(d, \mathit{Eng} - \mathit{Lndn} + \mathit{Paris})\right)$$



The analogy completion task

London : England :: Paris : ________

$$d^* = \text{argmax}_{d \in V} \left(\cos(d, \mathit{Eng} - \mathit{Lndn} + \mathit{Paris})\right)$$

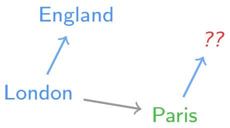


France
Switzerland
Italy
urology
swimming
purple
Latvia

The analogy completion task

London : England :: Paris : ________

$$d^* = \text{argmax}_{d \in V}\left(\cos(d, Eng - Lndn + Paris)\right)$$



Italy
Switzerland
France
urology
swimming
purple
Latvia

Unified Medical Language System (UMLS)

BMASS - BioMedical Analogic Similarity Set

Unified Medical Language System (UMLS)

Normalized concepts

C0009443 $\longrightarrow \left\{ \begin{array}{c} \text{common cold} \\ \text{cold} \\ \text{acute rhinitis} \end{array} \right\}$

BMASS - BioMedical Analogic Similarity Set

Unified Medical Language System (UMLS)

Normalized concepts

$$\text{C0009443} \longrightarrow \left\{ \begin{array}{c} \text{common cold} \\ \text{cold} \\ \text{acute rhinitis} \end{array} \right\}$$

Relation triples

$$\begin{array}{c} \textit{subject} \\ \textit{relation} \\ \textit{object} \end{array} \left\langle \begin{array}{c} \text{C0450297} \\ \text{RO:has-finding-site} \\ \text{C0341532} \end{array} \right\rangle$$

BMASS - BioMedical Analogic Similarity Set

ID	Name	Amb
Lab/Rx		
L1	form-of	1.0
L2	has-lab-number	1.1
L3	has-tradename	1.5
L4	tradename-of	1.3
L5	associated-substance	1.6
L6	has-free-acid-or-base-form	1.0
L7	has-salt-form	1.1
L8	measured-component-of	1.3
Hierarchical		
H1	refers-to	1.0
H2	same-type	10.4
Morphological		
M1	adjectival-form-of	1.1
M2	noun-form-of	1.0

ID	Name	Amb
Clinical		
C1	associated-with-malfunction-of-gene-product	2.6
C2	gene-product-malfunction-associated-with-disease	1.5
C3	causative-agent-of	4.6
C4	has-causative-agent	2.0
C5	has-finding-site	1.9
C6	associated-with	1.2
Anatomy		
A1	anatomic-structure-is-part-of	1.6
A2	anatomic-structure-has-part	5.4
A3	is-located-in	1.4
Biology		
B1	regulated-by	1.0
B2	regulates	1.0
B3	gene-encodes-product	1.1
B4	gene-product-encoded-by	2.4

Cross-product of 50 samples for each relation:

2,450 analogies for each relation $\Rightarrow$ **61,250 total analogies**

This dataset represents real biomedical relationships. . .

This dataset represents real biomedical relationships. . .

But it doesn't fit the standard paradigm!

"Unassuming" the standard assumptions

3 key assumptions in evaluation methodology:

- Single Answer
- Same Relationship(s)
- Informativity

Each is violated in recent analogy datasets

- Google[2], BATS[3], Sem-Para[4]

All are problematic in real-world data!

---

[2] Mikolov et al. 2013
[3] Gladkova et al. 2016
[4] Köper et al. 2015

"Unassuming" the standard assumptions

Single Answer | Same Relationship | Informativity

The given analogy has only one correct target.

- Enforced by argmax over candidates for completing the analogy.
- If multiple analogies, must get at least one wrong.

Problem cases

$$\text{flu} : \text{nausea} :: \text{fever} : \left\{ \begin{array}{c} \text{sweating} \\ \text{weakness} \end{array} \right\}$$

"Unassuming" the standard assumptions

Single Answer Same Relationship Informativity

The given analogy has only one correct target.

- Enforced by argmax over candidates for completing the analogy.
- If multiple analogies, must get at least one wrong.

Problem cases

$$\text{flu} : \text{nausea} :: \text{fever} : \left\{ \begin{array}{c} \text{sweating} \\ \text{weakness} \end{array} \right\}$$

Easy fix!

Allow for multiple correct answers; also report on all of them, for fuller picture.

"Unassuming" the standard assumptions

Single Answer | **Same Relationship** | Informativity

All information relating exemplars a and b also relates query c to d.

- Enforced by treating full vector difference as relation of interest.
- Many relations have partial overlap with one another.

Problem cases

"Unassuming" the standard assumptions

Single Answer | **Same Relationship** | Informativity

All information relating exemplars a and b also relates query c to d.

- Enforced by treating full vector difference as relation of interest.
- Many relations have partial overlap with one another.

Problem cases

$$\underbrace{\text{brother} : \text{sister}}_{\substack{\textbf{MaleCounterpart} \\ \text{SiblingOf}}} :: \underbrace{\text{husband} : \text{wife}}_{\substack{\textbf{MaleCounterpart} \\ \text{MarriedTo}}}$$

“Unassuming” the standard assumptions

Single Answer | **Same Relationship** | Informativity

All information relating exemplars a and b also relates query c to d.

- Enforced by treating full vector difference as relation of interest.
- Many relations have partial overlap with one another.

Problem cases

$\underbrace{\text{brother} : \text{sister}}_{\substack{\textbf{MaleCounterpart} \\ \text{SiblingOf}}} :: \underbrace{\text{husband} : \text{wife}}_{\substack{\textbf{MaleCounterpart} \\ \text{MarriedTo}}}$

$\underbrace{\text{wrist pain} : \text{shoulder pain}}_{\textbf{CoOccursWith}} :: \underbrace{\text{bipolar disorder} : \text{trazodone}}_{\substack{\textbf{CoOccursWith} \\ \text{TreatmentFor}}}$

"Unassuming" the standard assumptions

Single Answer | Same Relationship | **Informativity**

The relationship between exemplars a and b is specific enough to suggest the correct target d for query c.

- Issue with very broad semantic or hierarchical relationships.

Problem cases

"Unassuming" the standard assumptions

Single Answer | Same Relationship | **Informativity**

The relationship between exemplars a and b is specific enough to suggest the correct target d for query c.

- Issue with very broad semantic or hierarchical relationships.

Problem cases

Generally related *(UMLS)*

socks : stockings :: Finns : Finnish language

"Unassuming" the standard assumptions

Single Answer | Same Relationship | **Informativity**

The relationship between exemplars *a* and *b* is specific enough to suggest the correct target *d* for query *c*.

- Issue with very broad semantic or hierarchical relationships.

Problem cases

Generally related *(UMLS)*

socks : stockings :: Finns : Finnish language

Fix during dataset generation

Review samples from each relation to ensure they're properly determined.

Modifying the evaluation methodology

$$a : b :: c : \underline{\quad\quad} \quad \longrightarrow \quad d^* = \text{argmax}_{d \in V}\left(\cos(d, b - a + c)\right)$$

Evaluating analogies under 3 settings:

Modifying the evaluation methodology

$$a : b :: c : \underline{\quad\quad} \quad \longrightarrow \quad d^* = \mathrm{argmax}_{d \in V}\big(\cos(d, b - a + c)\big)$$

Evaluating analogies under 3 settings:

Single-Answer (SA) Single candidate target for each analogy selected as the only "correct" answer.

flu : nausea :: fever : $\left\{ \begin{array}{c} \textbf{sweating} \\ \cancel{\text{weakness}} \end{array} \right\}$ $\longrightarrow$

~~weakness~~
temperature
sweating
nodule

✗

Modifying the evaluation methodology

$$a : b :: c : \underline{\quad\quad} \quad \longrightarrow \quad d^* = \text{argmax}_{d \in V}\big(\cos(d, b - a + c)\big)$$

Evaluating analogies under 3 settings:

Single-Answer (SA) Single candidate target for each analogy selected as the only "correct" answer.

Multi-Answer (MA) All candidate targets for each analogy are considered to be correct.

flu : nausea :: fever : $\left\{\begin{array}{c}\text{sweating}\\ \text{weakness}\end{array}\right\}$ $\longrightarrow$

weakness
temperature
sweating
nodule

✓

Modifying the evaluation methodology

$$a : b :: c : \underline{\quad\quad} \quad \longrightarrow \quad d^* = \mathrm{argmax}_{d \in V}\big(\cos(d, b - a + c)\big)$$

Evaluating analogies under 3 settings:

Single-Answer (SA) Single candidate target for each analogy selected as the only “correct” answer.

Multi-Answer (MA) All candidate targets for each analogy are considered to be correct.

All-Info (AI) Use all possible exemplar objects and all candidate targets.

$$\text{flu} : \left\{ \begin{matrix} \text{nausea} \\ \text{cough} \end{matrix} \right\} :: \text{fever} : \left\{ \begin{matrix} \text{sweating} \\ \text{weakness} \end{matrix} \right\} \longrightarrow \begin{matrix} b - a \quad = \\ \frac{1}{2}\left(\begin{matrix} \text{nausea - flu} \\ + \text{ cough - flu} \end{matrix} \right) \end{matrix}$$

Evaluation methodology

Reporting 3 metrics over ranked candidates:

Acc_R Relaxed accuracy; correct if any valid answer is the top choice

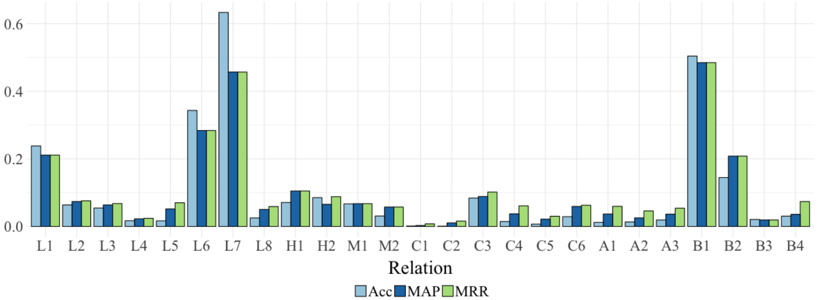
MAP Mean average precision

MRR Mean reciprocal rank

weakness
temperature
sweating
nodule

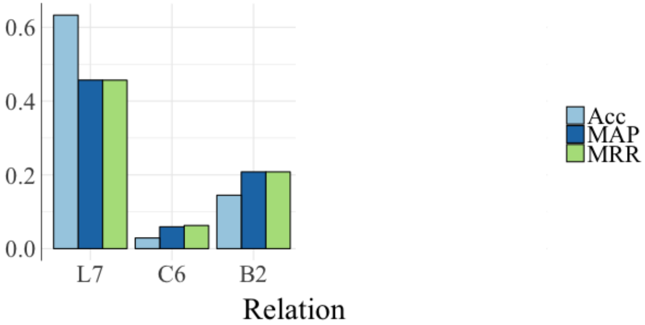
Acc_R	1.0
MAP	$\frac{5}{6}$
MRR	1.0

Overall results



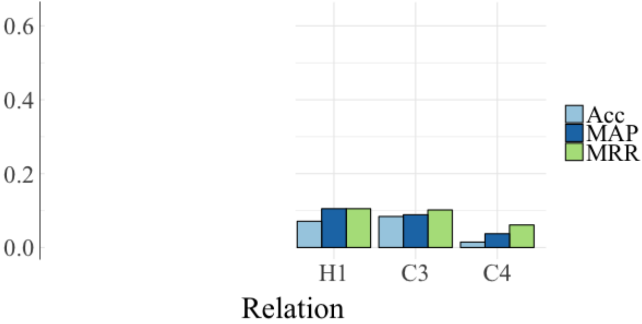
- Results shown for Multi-Answer setting.
- Average performance is around 11% on all metrics, with all embeddings. High variability between relationships!
- Used 5 different sets of embeddings trained on PubMed.

MAP/MRR give a better picture



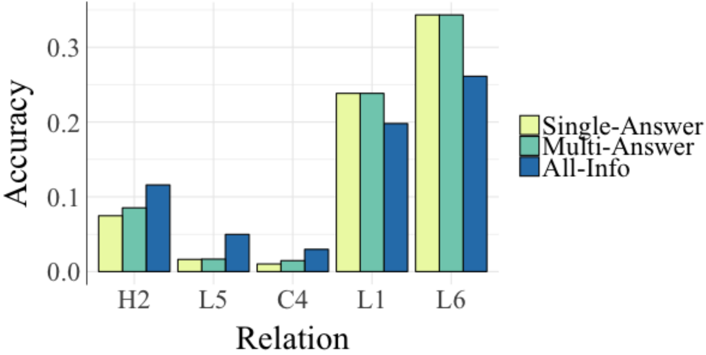
- MAP $<$ Acc_R indicates wider distribution of correct answers on *L7* (has-salt-form)
- MAP $>$ Acc_R shows that even if top answer is wrong, correct answers aren't far down on *C6* (associated-with), *B2* (regulates)

MAP/MRR give a better picture



- MRR $>$ Acc_R shows that the best correct answer stays near the top on *C4* (has-causative-agent)
- MRR $\approx$ Acc_R reflects more consistent positioning of nearest correct answer on *H1* (refers-to), *C3* (causative-agent-of)

All-Info benefits vary



- Extra examples help on *H2* (same-type), *L5* (associated-substance), and *C4* (has-causative-agent).
- But harm *L1* (form-of) (4% absolute) and *L6* (has-free-acid-or-base-form) (8% absolute)

Moving analogies forward

Single-Answer and Informativity assumptions addressed, but not Same Relationship(s).

Moving analogies forward

Single-Answer and Informativity assumptions addressed, but not Same Relationship(s).

- Drozd et al (2016) use a parametric logistic regression that can be used to learn affine subspaces.

Moving analogies forward

Single-Answer and Informativity assumptions addressed, but not Same Relationship(s).

- Drozd et al (2016) use a parametric logistic regression that can be used to learn affine subspaces.

Standard linear offset method does not work for real-world data!

- Our changes help, but overall performance is still low (as with other recent datasets). Use MAP and MRR!
- **Analogies are useful!** We need to find better ways to tackle this task.

Thank you!

Dataset and source code at:
https://www.github.com/OSU-slatelab/BMASS

newman-griffis.1@osu.edu