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
TL;DR

Increasing work on training embedding-based models for biomedical applications, but not many resources to evaluate on.
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](https://github.com/OSU-slatelab/BMASS)
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. 

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.

1https://github.com/OSU-slatelab/BMASS
The analogy completion task

BMASS

How can we make analogies more realistic?

Findings and challenges on our dataset
London : England :: Paris : France
The analogy completion task

London : England :: Paris :_________
The analogy completion task

London : England :: Paris : ________

\[ d^* = \operatorname{arg\,max}_{d \in V} \left( \cos(d, \text{Eng} - \text{Lndn} + \text{Paris}) \right) \]
The analogy completion task

London : England :: Paris : _______

\[ d^* = \arg\max_{d \in V} (\cos(d, \text{Eng} - \text{Lndn} + \text{Paris})) \]
The analogy completion task

London : England :: Paris : _____

\[ d^* = \arg\max_{d \in V} \cos(d, \text{Eng} - \text{Lndn} + \text{Paris}) \]

Italy
Switzerland
France
urology
swimming
purple
Latvia
Unified Medical Language System (UMLS)
Unified Medical Language System (UMLS)

Normalized concepts

C0009443 \rightarrow \{ common cold, cold, acute rhinitis \}
Unified Medical Language System (UMLS)

Normalized concepts

C0009443 \rightarrow \{ \text{common cold}, \text{cold}, \text{acute rhinitis} \}

Relation triples

subject relation object

C0450297

RO:has-finding-site

C0341532
### BMASS - BioMedical Analogic Similarity Set

#### Lab/Rx

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>form-of</td>
<td>1.0</td>
</tr>
<tr>
<td>L2</td>
<td>has-lab-number</td>
<td>1.1</td>
</tr>
<tr>
<td>L3</td>
<td>has-tradename</td>
<td>1.5</td>
</tr>
<tr>
<td>L4</td>
<td>tradename-of</td>
<td>1.3</td>
</tr>
<tr>
<td>L5</td>
<td>associated-substance</td>
<td>1.6</td>
</tr>
<tr>
<td>L6</td>
<td>has-free-acid-or-base-form</td>
<td>1.0</td>
</tr>
<tr>
<td>L7</td>
<td>has-salt-form</td>
<td>1.1</td>
</tr>
<tr>
<td>L8</td>
<td>measured-component-of</td>
<td>1.3</td>
</tr>
</tbody>
</table>

#### Hierarchical

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>refers-to</td>
<td>1.0</td>
</tr>
<tr>
<td>H2</td>
<td>same-type</td>
<td>10.4</td>
</tr>
</tbody>
</table>

#### Morphological

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>adjectival-form-of</td>
<td>1.1</td>
</tr>
<tr>
<td>M2</td>
<td>noun-form-of</td>
<td>1.0</td>
</tr>
</tbody>
</table>

#### Clinical

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>associated-with-malfunction-of-gene-product</td>
<td>2.6</td>
</tr>
<tr>
<td>C2</td>
<td>gene-product-malfunction-associated-with-disease</td>
<td>1.5</td>
</tr>
<tr>
<td>C3</td>
<td>causative-agent-of</td>
<td>4.6</td>
</tr>
<tr>
<td>C4</td>
<td>has-causative-agent</td>
<td>2.0</td>
</tr>
<tr>
<td>C5</td>
<td>has-finding-site</td>
<td>1.9</td>
</tr>
<tr>
<td>C6</td>
<td>associated-with</td>
<td>1.2</td>
</tr>
</tbody>
</table>

#### Anatomy

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>anatomic-structure-is-part-of</td>
<td>1.6</td>
</tr>
<tr>
<td>A2</td>
<td>anatomic-structure-has-part</td>
<td>5.4</td>
</tr>
<tr>
<td>A3</td>
<td>is-located-in</td>
<td>1.4</td>
</tr>
</tbody>
</table>

#### Biology

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>regulated-by</td>
<td>1.0</td>
</tr>
<tr>
<td>B2</td>
<td>regulates</td>
<td>1.0</td>
</tr>
<tr>
<td>B3</td>
<td>gene-encodes-product</td>
<td>1.1</td>
</tr>
<tr>
<td>B4</td>
<td>gene-product-encoded-by</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Cross-product of 50 samples for each relation:
2,450 analogies for each relation ⇒ **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\textsuperscript{2}, BATS\textsuperscript{3}, Sem-Para\textsuperscript{4}

All are problematic in real-world data!

\textsuperscript{2}Mikolov et al. 2013
\textsuperscript{3}Gladkova et al. 2016
\textsuperscript{4}Köper et al. 2015
“Unassuming” the standard assumptions

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 ________

flu : nausea :: fever : \{ sweating, weakness \}
“Unassuming” the standard assumptions

Single Answer

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 ---

flu : nausea :: fever :

\{ sweating, weakness \}

Easy fix!

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

\[
\begin{align*}
\text{brother} : \text{sister} &::\quad \text{husband} : \text{wife} \\
\underbrace{\text{MaleCounterpart}}_{\text{SiblingOf}} &::\quad \underbrace{\text{MaleCounterpart}}_{\text{MarriedTo}}
\end{align*}
\]
“Unassuming” the standard assumptions

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**

\[
\begin{align*}
\text{brother} : \text{sister} &:: \text{husband} : \text{wife} \\
\text{MaleCounterpart} &\quad \text{SiblingOf} \\
\text{MaleCounterpart} &\quad \text{MarriedTo} \\
\text{wrist pain} : \text{shoulder pain} &:: \text{bipolar disorder} : \text{trazodone} \\
\text{CoOccursWith} &\quad \text{CoOccursWith} \\
\text{TreatmentFor} &\quad \\
\end{align*}
\]
“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 ---
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 (\textit{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

\[ a : b :: c : ____ \rightarrow \quad d^* = \arg\max_{d \in V}(\cos(d, b - a + c)) \]

Evaluating analogies under 3 settings:
Modifying the evaluation methodology

\[ a : b :: c : \quad \rightarrow \quad d^* = \arg\max_{d \in V} (\cos(d, b - a + c)) \]

Evaluating analogies under 3 settings:

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

\[
\begin{align*}
\text{flu} : \text{nausea} \quad &:: \quad \text{fever} : \\
&\quad \{ \text{sweating, weakness} \}
\end{align*}
\]

weakness

| temperature  | sweating | nodule |
Modifying the evaluation methodology

\[ a : b :: c : \quad \rightarrow \quad d^* = \arg\max_{d \in V} (\cos(d, b - a + c)) \]

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 :
  - sweating
  - weakness

  weakness
temperature
sweating
node
Modifying the evaluation methodology

\[ a : b :: c : \quad \rightarrow \quad d^* = \arg\max_{d \in V} (\cos(d, b - a + c)) \]

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.

\[ \text{flu} : \{ \text{nausea, cough} \} :: \text{fever} : \{ \text{sweating, weakness} \} \quad \rightarrow \quad \frac{1}{2} (\text{nausea} - \text{flu} + \text{cough} - \text{flu}) \]
Evaluation methodology

Reporting 3 metrics over ranked candidates:

- $\text{Acc}_R$: Relaxed accuracy; correct if any valid answer is the top choice
- MAP: Mean average precision
- MRR: Mean reciprocal rank

weakness
temperature
sweating
nodule

$\text{Acc}_R = 1.0$
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

- \( \text{MRR} > \text{Acc}_R \) shows that the best correct answer stays near the top on \( C4 \) (has-causative-agent)
- \( \text{MRR} \approx \text{Acc}_R \) 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