Learning to Describe Unknown Phrases with Local and Global Contexts

Shonosuke Ishiwatari¹, Hiroaki Hayashi², Naoki Yoshinaga¹, Graham Neubig²
Shoetsu Sato¹, Masashi Toyoda¹, and Masaru Kitsuregawa¹

¹The University of Tokyo
²Carnegie Mellon University Language Technologies Institute
... I was in 3rd grade during the **sonic boom** tour; one...

**sonic boom** | ˈsænik bɔːm | noun

a loud explosive noise caused by the shock wave from an aircraft traveling faster than the speed of sound.
How do human infer the meanings?

STEP1: Search web to find other usages
- “sonic boom”
  - that a **sonic boom** happens...
  - the **sonic boom** kick-off concert...
  - planes cause **sonic boom** when...
  - albums are “**sonic boom**” and...
  - for the **sonic boom** bonus disc...

STEP2: Classify the usages
- **Related to natural phenomenon?**
- **Related to music?**
  - that a **sonic boom** happens...
  - the **sonic boom** kick-off concert...
  - planes cause **sonic boom** when...
  - albums are “**sonic boom**” and...
  - for the **sonic boom** bonus disc...
How do human infer the meanings?

**STEP 1:** Search web to find other usages

- "sonic boom" search results:
  - that a **sonic boom** happens...
  - the **sonic boom** kick-off concert...
  - planes cause **sonic boom** when...
  - albums are "**sonic boom**" and...
  - for the **sonic boom** bonus disc...

**STEP 2:** Classify the usages

- that a **sonic boom** happens...
- the **sonic boom** kick-off concert...
- planes cause **sonic boom** when...
- albums are "**sonic boom**" and...
- for the **sonic boom** bonus disc...

**STEP 3:** Interpret them

- I was in 3rd grade during the **sonic boom** tour; one...

- An album by KISS?
RQ: Can machines help us infer the meanings?

**STEP 1:** Map the usages to vector space

... that a **sonic boom** happens ...
... the **sonic boom** kick-off concert ...
... planes cause **sonic boom** when ...
... albums are “**sonic boom**” and ...
... for the **sonic boom** bonus disc ...

**STEP 2:** Describe the target phrase reading the context

**Input:** ... I was in 3rd grade **during the** **sonicBoom** tour; one...

**Output:** Album by KISS.
1. Context-aware description generation task
   - Large-scale dataset: 989k entries
   - Multi-domain: name of places, movies, persons, ...
   - Various expressions: phrases as well as words

2. Description generation model which
   - achieves SOTA performances on 4 datasets
   - combines local & global context dynamically

... the sonic_boom tour ...
## Task: Context-aware description generation

**WPDD: Wikipedia Phrase Description Dataset**

| Phrase:       | sonic boom |
|---------------|------------|
| **Context:**  | This tour is the first official tour to support the band's latest studio effort, 2009's [TRG]. |
| **Description:** | album by Kiss |

Can we make this dataset automatically?
Article: *Kiss Sonic Boom Over Europe*

... This tour is the first official tour to support the band's latest studio effort, 2009's *Sonic Boom*. ...

**Extracted Data**

**Phrase:** sonic boom

**Context:** This tour is the first official tour to support the band's latest studio effort, 2009's *Sonic Boom*. ...

**Description:** album by Kiss

**Item:** *Sonic Boom* (Q908201)

- **Description:** album by Kiss
- **Instance of:** album
- **Genre:** hard rock
1. Context-aware description generation task
   - **Large-scale**: 989k entries
   - **Multi-domain**: name of places, movies, persons, ...
   - **Various expressions**: phrases as well as words

2. Description generation model which
   - achieves **SOTA** performances on 4 datasets
   - combines **local** & **global context** dynamically

... the *sonic_boom* tour ...
Baseline: Definition-generation [Noraset+ 17]

- Seed-feeding to condition LSTM language model

Definition of the word “lira” in a dictionary

The input embedding loses its influence as the definition becomes longer
Baseline: Definition-generation [Noraset+ 17]

- Seed-feeding + Gate function

Controls the influence of the word embedding
Baseline: Definition-generation [Noraset+ 17]

Seed-feeding + Gate function

 Controls the influence of the word embedding
Baseline: Definition-generation [Noraset+ 17]

- Seed-feeding + Gate function

Controls the influence of the word embedding
Seed-feeding + Gate function

The input information is fixed regardless of context

Baseline: Definition-generation [Noraset+ 17]
Proposed: Local & Global Context-aware model

Global context

- CBOW
  - sonic boom
- CNN
  - Character info. [Noraset+ 17]

Description decoder

- LSTM
  - $S_t$
- GATE
  - $S'_t$
- LSTM
  - $S_{t+1}$

moving
Proposed: Local & Global Context-aware model

Local context [Ni+ 18]

... caused by [TRG] or by ...

Global context

sonic boom

Attention

sonic boom

Description decoder

LSTM

GATE

S_{t+1}
Proposed: Local & Global Context-aware model

**Local context [Ni+ 18]**

\[ \ldots \text{caused by [TRG]} \quad \text{or by} \quad \ldots \]

**Global context**

\[
\begin{align*}
\text{sonic boom} & \rightarrow \text{CBOW} \\
\text{sonic boom} & \rightarrow \text{boom} \\
\text{CNN} & \rightarrow \text{Transformer} \\
\end{align*}
\]

**Attention**

\[ \text{sonic boom} \]

**Character info. [Noraset+ 17]**

\[ s o n i c \_ b o o m \]

**Global cont.:** general meaning in large corpus

**Local cont.:** specific usage in a sentence

\[ \ldots \text{caused by or by} \ldots \]
Experiments: Can the model utilize both contexts?

Dataset
- WPDD (Proposed)
- WordNet [Noraset+ AAAI17]
- Oxford Dictionary [Gadetsky+ ACL18]
- Urban Dictionary [Ni+ IJCNLP17]

Methods
- **LOG-CaD (Proposed)**: Local & Global Context-aware Desc. Gen.
- **Global**: Uses word embedding as context [Noraset+ 17]
- **Local**: Uses local context as context [Ni+ 17]

Evaluation
- Automatic (BLEU) & Manual (5-levels, 3 En. speakers) eval.
**Result: SOTA on four datasets**

100% of the words have global context (i.e., word embeddings)

Only 27% of the words have global context

| BLEU ↑  | WordNet  | Oxford  | Urban  | Wikipedia |
|---------|----------|---------|--------|-----------|
| **Global** | 24.10    | 15.05   | 6.05   | 44.77     |
| **Local**  | 22.34    | 17.90   | 9.03   | 52.94     |
| **Proposed** | 24.79 | 18.53   | 10.55  | 53.85     |

Local cont. is essential if global context is unavailable
Example

Input: “daniel o’ neil”

| Cont. #1 | after being enlarged by publisher daniel o’neill it was reportedly one of the largest and most prosperous newspapers in the united states. |
| Cont. #2 | in 1967 he returned to belfast where he met fellow belfast artist daniel o’neill. |

| Method     | Output #1       | Output #2       |
|------------|-----------------|-----------------|
| Global     | american musician | american musician |
| Local      | american publisher | british musician |
| Proposed   | american writer  | british musician |
| Reference  | american journalist | Irish artist |

Same outputs for different contexts 🤔
## Result: Manual evaluation

| Score   | Local | Proposed |
|---------|-------|----------|
|         | 2.7   | 3.0      |

### Annotated Scores:
- 1: completely wrong
- 2: correct topic w/ wrong info.
- 3: correct but incomplete
- 4: small details are missing
- 5: correct

### Input: “auto race”

| Context | the regulating body for pari-mutuel keirin and **auto race** competitions in Japan. |
|---------|------------------------------------------------------------------------------------|
| Local   | a type of sports equipment                                                         |
| Proposed| type of motor racing                                                               |

Averaged score:
- **Local**: 1.3/5
- **Proposed**: 4.0/5
How do the models utilize local context?

Local context provides more gain if global context is unreliable.
How do the models utilize global context?

- Local context is extremely short
- Local context is long

Global context is essential if the local context is not informative.

[Graph showing BLEU scores against the length of local context (in words), with proposed and local context lines, and annotations for +1.7 and +0.4.]
Summary

- Research Question
  - Can machines help us describe unknown phrases?
- Answer to RQ
  - Yes. In average, the generated descriptions are correct but incomplete
- Contributions
  - Large, multi-domain phrase description dataset
  - SOTA model that captures local & global context
- Future work
  - Use multiple local contexts to describe phrases