SEAS RESEARCH & DEVELOPMENT SHOWCASE

Automatic Generation of Context-Based Fill-in-the-Blank Exercises Using Vector Space Models and Google *n*-grams

Jennifer Hill (Advisor: Rahul Simha)

INTRODUCTION

According to the American Library Association, 14% of adults in the United States cannot "search, comprehend, and use continuous texts" [1]. Current government and philanthropic funding only indirectly helps one tenth of these nearly 30 million individuals.

To address this shortcoming, we aim to automatically create reading comprehension exercises from existing text passages. We specify that a successful comprehension exercise should challenge a reader's contextual understanding of the passage's meaning rather than his or her vocabulary understanding. We propose a method of automatically generating fill-in-the-blank exercises to challenge and improve comprehension skills, using a unique application of word co-occurrences vector space models and the Google *n*-grams database.

CONTEXTUAL AWARENESS

We propose that a good reading comprehension question challenges the reader not with *syntactic* errors or unusual word choices, but with *contextual inconsistencies*.

We specify that a good distractor should make sense grammatically and conceptually within a narrow context, but should not make sense within the broader context of the entire sentence.

to stay _____ during	***Follow these tips to stay _____ during a hurricane.***
● open ● safe ● quiet ● active	○ open ● safe ○ quiet ○ active

Example showing distractor applicability in a narrow vs. full context.

When looking at a narrow context, all four words fit in the blank, but when the meaning implied by the entire sentence is taken into account, only one makes sense. Thus, a reader must be actively constructing meaning from the sentence as they read rather than simply decoding the individual words.

CHOOSING BLANKS

We choose to blank out words that have strong contextual links to words in the surrounding text, leaving enough context for the reader to understand the sentence's intended meaning when that word is removed.

To determine contextually-linked words, we utilize word co-occurrence likelihoods from the word vector space model GloVe [2]. We assume that words that are paired together regularly are likely to have a notable contextual and semantic relationship.

```
ts = target sentence //minus stop words + proper nouns
ss = scope sentences //minus stop words

M_ss = matrix of word vectors in scope sentences

for word w_ts in ts:
   if count(w_ts) == 1:
      V_wts = vector representation of w_ts
      dist = V_wts · M_ss //cosine distance
      w_p = word in ss with smallest dist
      closest[w_ts] = dist[w_ts][w_p]
```

Algorithm for selecting blanked words by finding closest co-occurring pairs.

We adjust the contextual "scope" to allow us to incorporate potentially relevant information from previously-read text which can contribute to the understanding of the current sentence. We test scopes containing just the current sentence, and those containing 1 and 2 prior sentences.

EVALUATION

Our corpus contained 18 passages obtained from ReadWorks.org (Lexile Level 100 to 1000). For each passage, we generated fill-in-the-blank questions for each scope, resulting in 170 unique questions.

53 human volunteers answered an anonymous questionnaire with both blanked phrases and full passage exercises.

We evaluate our questions on their validity and reliability [4]. A question is considered **valid** if it tests a user's understanding of the given text, and **reliable** if only one answer can reasonably fill the given blank.

CHOOSING DISTRACTORS

We explore a unique application of the Google *n*-grams corpus [3] for generating distractors for our fill-in-the-blank questions. We find all words with the same part of speech as the blanked word found in the Google *n*-grams database.

```
- replace w_blank with POS
- replace pronouns with POS
- replace person entities with POS
- for each subtree containing w_blank:
   - get all 2- to 5-grams containing w_blank

- select all Google n-grams matching pattern
- distractors = all words at index i[w_blank]
- remove synonyms of w_blank
```

Algorithm for selecting distractors by finding matching Google n-grams.

EXAMPLE

"It's written for and put together by the fifth graders," Dr. Reed said.
"It's written for and put [RR] by the fifth graders," [PRP] said.

ngrams	words at index $i[w_{blank}]$
and put [RR]	forward, down, forth, off, here, ...
put [RR] by the	up, out, away, back, off, aside, ...
[RR] by the fifth graders	✖
→ [RR] by the [persons]	up, down, off, directly, only, ...

Validity

- **Blanks**: rate question on a 1-5 quality scale (without distractors)
- **Distractors**: examine ratio of words that fit blank in narrow context to words that fit broader context.
  - Narrow: Correct 100%, Incorrect 100%
  - Full: Correct 100%, Incorrect 0%

Reliability

- **Distractors**: examine percentage of correct answers selected when given the full context

RESULTS & CONCLUSIONS

BLANKS

Rating	Percent of Responses		
5	49.7%	75.3%	GOOD
4	25.6%		
3	15.9%		
2	6.5%	8.8%	BAD
1	2.3%		

Table 1: Distribution of ratings for all generated questions

DISTRACTORS

Scope	Target	Distractors	Target	Distractors
1 sentence	90.9%	65.4%	98.1%	9.7%
2 sentences	86.7%	62.8%	97.1%	9.8%
3 sentences	88.1%	62.8%	96.1%	11.7%

Table 2: Percentage of target and distractor words determined to fit each blank

Scope	Correct Answers
1 sentence	90.27%
2 sentences	90.23%
3 sentences	88.33%

Table 3: Percentage of questions answered correctly

Preliminary results suggest that our algorithms are effective at both selecting blanks and generating distractors when automatically creating exercises to test reading comprehension. A single-sentence scope seems the most effective for finding contextually-linked words.

These findings suggest a promising future for the automatic generation of literacy-based exercises.

REFERENCES

[1] Baer, Kutner, Sabatini, and White. Basic Reading Skills and the Literacy of America's Least Literate Adults: Results from the 2003 National Assessment of Adult Literacy (NAAL) Supplemental Studies. *National Center for Education Statistics.* 2009.

[2] Pennington, Socher, and Manning. GloVe: Global Vectors for Word Representation. 2014.

[3] Michel, Shen, Aiden, Veres, Gray, Brockman, Pickett, Hoiberg, Clancy, Norvig, Orwant, Pinker, Nowak, and Aiden. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science.* 2010.

[4] Alderson, Clapham, and Wall. Language Test Construction and Evaluation. Cambridge University Press. 1995.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

DEPARTMENT OF
COMPUTER SCIENCE
SCHOOL OF ENGINEERING & APPLIED SCIENCE