WORD SENSE DISAMBIGUATION AND SEMANTIC PARSING MODELS ARE ADVANCED EDUCATORS

Chan Woo Kim
LXPER Inc.
Williams College
ck15@williams.edu

September 11, 2021

ABSTRACT

This paper proposes a novel method of generating standard but deeper types of questions that test a student's true understanding of certain words. Using Facebook's Gloss-Informed Bi-encoder Word Sense Disambiguation (WSD) [2] model and the semantic parser in spaCy's Natural Language Processing model [4], we have designed novel pipelines that could produce three incredibly deep and nuanced vocabulary questions that test a student's deeper understanding over a vocabulary set. These include the following types of questions: "Words in Context", "Best Use of a Word", and "Fill-in-the-Blank".

1 Introduction

Obtaining complete mastery over a new vocabulary set requires something additional than merely memorizing the definitions. It requires a nuanced understanding of the context in which the words can be used and understood, and the many different meanings they possess (especially in the case of homonyms). However, in Korea and in many vocabulary learning applications, often the standard way for students to study a vocabulary set is to mindlessly memorize and regurgitate the translations/definitions of these words. Surely they may be provided with example sentences and related words, but *automatically generating* a test that truly evaluate these deeper and more nuanced concepts to students is extremely rare. For instance, the testing feature on Quizlet asks students to match the words in the study set to their corresponding definitions, and that is indeed the standard model for automatic testing. This is mainly because there previously was not a clear approach to automatically generate questions that are more meaningful than definition matching.

Though we can not ignore the importance of memorizing the definition in the first place, there is clearly a need for a deeper way to test one's understanding. For instance, in order to solve a question like that of TABLE 1, the student must acquire a completely comprehensive understanding of the words involved and beyond. Our contribution is the development of novel pipelines that can automatically generate such questions only given a word like "directly" in the case of the example in TABLE 1.

The implications are clear; students tend to learn incredibly well by guided trial and error through being tested. By being able to automatically generate more difficult and nuanced questions, we are able to bring about more sophisticated level of vocabulary internalization from the students.

2 Models Involved

This section describes the models found in literature that were used in the construction of our technology.

2.1 Gloss-Informed Bi-encoder Word Sense Disambiguation (WSD)

The function of a WSD model is to detect the right meaning of a word given a context. For instance, words like "model" are homonyms; it can refer to an attractive person who wears clothes to display fashions, a "hypothetical description of a complex entity or process", and more, according to WordNet [5]. The WSD model is then tasked with detecting which meaning is used in a given use of the word in a sentence.

The Gloss-Informed Bi-encoder uses two instances of deep transformer encoder networks that are initialized as BERT models. This paper will not discuss in specific details what transformers and the BERT pre-trained models are, but they can be found here: [6][3].

A typical task for a WSD model is the following: it is given a sentence like the following (taken from Word Hippo [1]):

"After some hours of intense work, we had macheted a path through the jungle to the bank of the river."

then it is tasked with finding the correct definition of words such as "bank" in this context.

In order to understand what the Gloss-informed WSD model does, we must know that BERT encoders have the property that the input size and the output size are the same.

For instance, let c be the input sentence, $c = [c_0, c_1, ..., c_w, ..., c_n]$, where c_w is our "target word". For the above example,

$$c = [\text{"After", "some", ..., "bank", "of", "the", "river"}],$$

where $c_w =$ "bank".

Let $BC(c) = z$ be the BERT encoder that encodes sentences like c shown above to a new representation z as its output. Specifically, each $c_i \in c$ is encoded by vector $z_i \in z$.

In other words:

$$BC(c) = z = [z_0, z_1, ..., z_w, ..., z_n]$$

where z_w is the deep, semantic, and context-based *encoding* vector of c_w. In this example, z_w is a semantic representation of the word "bank".

Now let $BC(c)$ be the BERT model used for encoding the *input sentence* and let $T(g)$ be the BERT model used for encoding *definitions found in WordNet*.

Given these tools, Gloss-Informed WSD model ultimately does the following:

1) Encode the sentence c using BERT to z: $BC(c) = z = [z_0, z_1, ..., z_w, ..., z_n]$.
2) Find all definitions of "bank" found in WordNet [5], all 17 of them: $g = [g_0, g_1, ..., g_{16}] =$ ["sloping land (especially the slope beside a body of water)", "a financial institution that accepts deposits and channels the money into lending activities", ...].
3) Encode all the above definitions using the BERT model T. Get a list of definition encodings like so:

$$[T(g_0), T(g_1), ..., T(g_{16})]$$

Understanding that $g_i = [g_{i_1}, g_{i_2}, ..., g_{i_n}]$, or in the above example:

$$g_0 = [\text{"[CLS]", "sloping", "land", "(", "especially", "the"}]$$

Since BERT has the property that the input size and the output sizes are the same,

$T(g_0)[0] =$ encoding of "[CLS]"
$T(g_0)[1] =$ encoding of "sloping"
$T(g_0)[2] =$ encoding of "land"

and so on.

For this purpose, we let $r_{s0} = T(g_0)[0]$ be the global representation of g_0. The entirety of the definition corresponding to g_0 is *represented* in the vector defined as r_{s0}.

Question	Given the context, what is the closest meaning of the word "directly"?
Sentence	"Akira came **directly**, breaking all tradition. Was that it? Had he followed form... I ask **directly** because the use of a go-between takes much time."
Choices	a) Frankly b) Without Mediation c) Confidently d) With Precision
Answer	b) Without Mediation

Table 1: Example "Words in Context" Question From The SAT Practice Exam 1

Question	Given the context, what is the closest meaning of the word **"temper"**?
Sentence	"He had a wild **temper**, and when sufficiently enraged could suffer seizures and blackouts."
Choices	a) A sudden outburst of anger b) The elasticity and hardness of a metal object; its ability to absorb considerable energy before cracking c) A characteristic (habitual or relatively temporary) state of feeling d) A disposition to exhibit uncontrolled anger
Answer	d) A disposition to exhibit uncontrolled anger

Table 2: Example "Words in Context" Generated By Our Model

Now we end up with the following deep representations::

z_w, the context-based semantic encoding of the word "bank" and

$$r = [r_{s0}, r_{s1}, ..., r_{s16}],$$

the list of semantic encodings of the possible definitions of the word "bank".

One by one, we calculate a score:

$$\phi(z_w, r_{s_i}) = z_w \cdot r_{s_i}$$

for $i = 0, 1, 2, ..., 16$. Then the r_{s_i} that obtains the highest score is chosen. Since r_{s_i} is the semantic encoding of the i^{th} definition of "bank", we then return the definition corresponding to r_{s_i} and deem that to be the correct definition of our target word in the context.

2.2 spaCy's Semantic Parsing Model

3 Types of Questions

The following introduces the types of questions that our generation pipelines can automatically create. Before delving into our actual models, this section simply introduces today's standards when it comes to these question types by displaying questions from publicly available exams.

The key philosophies that guided the choice of these questions and the design of our engines are that: 1) the questions must be utterly unsolvable without at least knowing the definitions and 2) the question must be challenging even to those that know and have used the word. The idea is to be something more than a definition matching testing software; we had concluded from user interviews that one could choose the right definition given a test without even knowing what it means. This is because the act of "recognition" that is required for matching definitions does not require even the slightest of true internalization of its meaning. For this reason, we've specifically designed automatic testing engines that generate questions users can not get away with without completely understanding the vocabulary that's involved.

3.1 Words in Context

One such example of "Words in Context" questions can be seen at TABLE 1. The advanced idea of questions like this is that they require something entirely additional to knowing the simple definitions of words such as "frankly", "mediation", "precision", and so on. One must have full mastery over context understanding and the full spectrum of meaning involved with the above list of words.

A key requirement for this level of question is that the incorrect answers are all technically correct; in this example, the wrong answers could technically replace the word "directly" in other contexts and even in the given sentence without sounding awkward. However, the "closest meaning" is clearly option "b", given the contextual clue that otherwise "a go-between takes much time".

TABLE 2 displays the output of our model, only given the word "temper". The nuance of this question is that all the given options are correct definitions of "temper" and "a sudden outburst of anger" is a very common way people define the word "temper". But the context uses the word as something describing the personality of the subject, and after considering options "c" and "d", we hope that students would uncover enough context clues and have enough intuition about the English language to finally choose "d".

3.2 Best Use of Word

This is perhaps a less common type of vocabulary question but we nevertheless believe that this follows the design philosophy explained above.

3.3 Fill In The Blank

4 Model Pipelines

4.1 Words in Context

The diagram representation of the pipeline that generates Words in Context questions can be found in Figure 1.

4.1.1 Patent Ingenuity Description

For the patent, ***the ingenuity of this pipeline*** is the cumulative application of two irrelevant systems, the Example Sentence API and the WSD model, combined with our custom "Softmax-to-WordNet Options" algorithm, that come together to generate incredibly advanced test questions. To elaborate further, the fundamental purpose of the WSD model is to detect the one correct meaning of each word in a sentence and it was never meant to combine with other modules to generate test questions. In other words, the ingenuity of this pipeline is that we directly manipulated the last layers of this model with a custom algorithm and combined it with other systems such that we've developed an original way to automatically generate meaningful test questions like the one in Table 2.

With just the information that our user is studying the word "temper", the pipeline automatically generates questions like that in Table 2. First we pull an example sentence that uses "temper" from the web. Then we use the WSD model to find the correct definition of "temper" given that context sentence. Finally, we find the distraction answer choices:

4.1.2 Distraction Choosing Algorithm: "Softmax-to-WordNet Options"

But the selection of distraction answer choices is the challenge of this model. Recall the score function ϕ defined at the end of **Section 2.1**. Instead of simply choosing the definition with the highest score, we make a sorted list of all the scores, and consider the list excluding the highest score as our potential distraction options.

However, especially since these models were never designed to focus into the distraction options, the algorithm additionally aims to solve the following three problems:

Problem 1. In some cases, multiple definitions found in WordNet work all perfectly fine in describing the meaning of a word in context. For instance, for the word "model", there are these two definitions in WordNet among other: "a hypothetical description of a complex entity or process" and "representation of something (sometimes on a smaller scale)". For a sentence like: "In front of you is a model of the Sun and the chemical processes that produce its heat", both definitions are both too correct for one to be deemed objectively incorrect.

Problem 2. In some cases, a word only has one or two definitions such that we are not able to create all four answer choices. For instance, the word "serendipity" and "auspicious" both only have one definition in WordNet. Therefore,

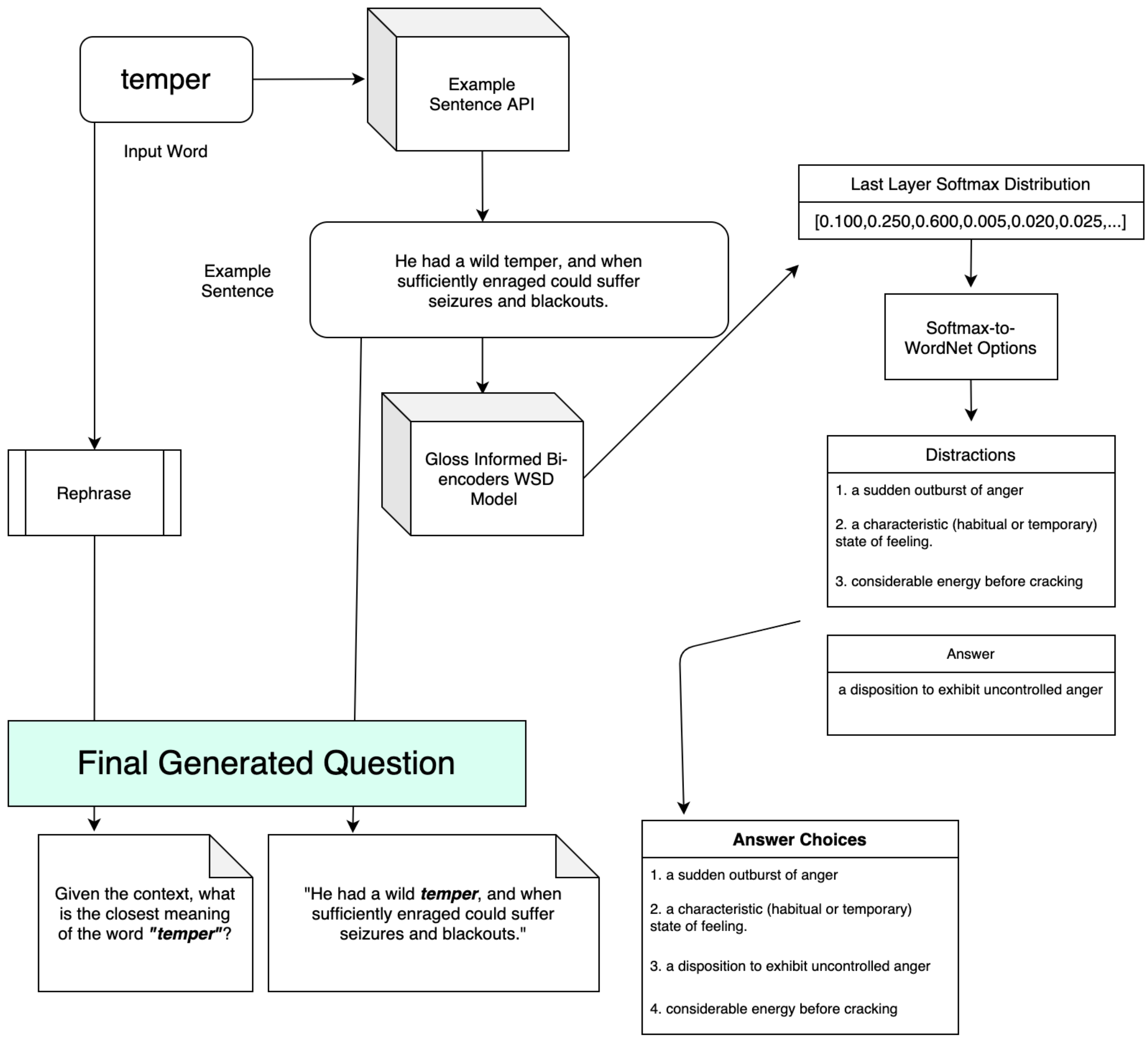

Figure 1: "Words in Context" Generation Pipeline Diagram

that one definition will always be the "answer" and we do not have other definitions to consider to make up the rest of the answer choices.

Problem 3 A vast majority of words in the English language possess multiple parts of speech. The word "model" can be used both as a verb and a noun. Preliminary user demos have shown that it is undesirable for any of the answer choices to describe the meaning in a different part of speech or tense as how it is used in the sentence. For instance, we want to remove instances like "a sudden outburst of anger" being an answer choice when the given sentence is "I love tempering chocolate when making my milk chocolates", since the former describes a noun while "tempering" in the sentence is used as a verb.

In order to solve the above two issues, we apply the following algorithm.

1. Retrieve all WordNet definitions of the input word.
2. If those number of definitions is less than 20, then find words synonymous to the input word.
3. Retrieve all WordNet definitions of each synonym.
4. Finalize a list of 20 items consisting of WordNet definitions of input word + WordNet definitions of its synonyms. This solves **Problem 2** since we end up with definitions that are close enough in meaning to the

input word for the question to be challenging. If we solved this issue by finding random definitions, then we'd be left with questions that are too obvious. In order to solve **Problem 3**, we do a simple Part of Speech filtering to only have 20 items describing the same part of speech.

5. Run WSD model except the last layer and sort the outputs to get a sorted list of scores for each of the above 20 definitions. The definition with the highest score is the "answer". Now we must find the distraction options from the list of 19 items.
6. In order to solve **Problem 1**, we apply a threshold for weight differences. Specifically, if the difference between two scores are too low, they are considered too close in meaning and the latter is rejected.
7. Continue rejecting based on the above criteria until we are left with three "nice" distraction options from the above 19. Though it is indeed possible that all 19 get rejected, but that has been empirically shown to be near impossible through testing. But for the extremest of events, we've added a final check to loop back to **step 2** again if that ever occurs.

4.2 Best Use of Word

4.3 Fill In The Blank

References

[1] Kat IP Pty Ltd 2008. Word Hippo. Retrieved from https://www.wordhippo.com/.

[2] Terra Blevins and Luke Zettlemoyer. Moving down the long tail of word sense disambiguation with gloss-informed biencoders. *arXiv preprint arXiv:2005.02590*, 2020.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[4] Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.

[5] George A Miller. *WordNet: An electronic lexical database*. MIT press, 1998.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

5 Appendix

5.1 Example Generated Questions

1. ____________ Given the context, what is the closest meaning of the word "minute"?

We were now encountering the last minute and Dingle had all hands on deck to rescue their rapidly sinking ship.

a) an indefinitely short time

b) a particular point in time

c) a unit of angular distance equal to a 60th of a degree

d) a unit of time equal to 60 seconds

2. ____________ Given the context, what is the closest meaning of the word "issue"?

As Palestinians saw it, at issue was a policy of ignorization which aimed at destroying their cultures academic capacities.

a) one of a series published periodically

b) an important question that is in dispute and must be settled

c) some situation or event that is thought about

d) a phenomenon that follows and is caused by some previous phenomenon

3. ____________ Given the context, what is the closest meaning of the word "engage"?

Furthermore, in some respects Wiccans regularly engage in practices against which theurgists themselves warned, except in very exceptional circumstances.

a) ask to represent; of legal counsel

b) give to in marriage

c) consume all of one's attention or time

d) carry out or participate in an activity; be involved in

4. ____________ Given the context, what is the closest meaning of the word "straight"?

Much of western architecture emphasizes repetition of simple motifs, straight lines and expansive, undecorated planes.

a) a heterosexual person; someone having a sexual orientation to persons of the opposite sex.

b) a poker hand with 5 consecutive cards (regardless of suit)

c) free from curves or angles

d) having no deviations

5. ____________ Given the context, what is the closest meaning of the word "level"?

Licentiate degrees vary widely in their meaning, and in a few countries are doctoral level qualifications.

a) a position on a scale of intensity or amount or quality

b) a specific identifiable position in a continuum or series or especially in a process

c) a relative position or degree of value in a graded group

d) height above ground