Teaching Culturally and Linguistically Diverse International Students in Open and/or Online Learning Environments: A Research Symposium

On the Role of Machine Learning in A Human Learning Process

Peng Wang
School of Translation and Interpretation
University of Ottawa, Ottawa, Ontario
Email: pwang2@uottawa.ca

Abstract

This study aims to explore the role of machine learning in a human learning process. In particular, we use word embedding and basic recurrent neural network methodologies in the teaching of terminology and specialized knowledge acquisition for translation students. Results show that word distribution in vector space trained on a relevant corpus can provide useful insights for learners to understand the terms and their associated concepts. We also build term recognition models for different levels of learners, which help instructors predict terms for these learners, while incorporating their previous knowledge and skills, so as to better communicate with the students in their teaching.

Keywords: machine learning (ML), human learning, terminological schematic context, word embedding

Introduction

A Paradigm Shift from Teaching to Learning

To some degree, the COVID-19 pandemic has revolutionized the teacher-student relationship in higher education. A more radical transition, from traditional face-to-face to online teaching after 2020, has ushered in a new era of digitized education, catalyzing the already growing shift in the instructor's role. In a virtual classroom, instructors are more content facilitators than content providers, using the Web as the primary teacher-student link in a teaching environment, without much visual control provided by direct eye contact (Fein and Logan 2003). From a communicative perspective, a face-to-face teaching environment presumably has much more diverse inputs for teacher-student interaction, whereas online teaching mostly highlights one or more particular features, such as voice and facial expressions.

This relatively restrained, accentuated online environment can present new challenges and opportunities for instructors. For example, Kebritchi et al. (2017) summarized three major categories of issues for teaching successful online courses in higher education, namely those related to online learners, instructors, and content development. Choi and Park (2006) pointed out that novice instructors find that online courses involve a heavy workload, technology issues, and student-teacher interaction. It is necessary for instructors to innovate their pedagogical methods and frameworks to meet these challenges.

Further, in an online classroom, an instructor has transitioned into a new role as a learning facilitator, guiding and assisting students in learning for themselves, rather than a knowledge transmission agent. As Kebritchi et al. (2017) argued, online courses place emphasis on the ability to deliver content, which is a transfer from teacher-centered to student-centered education; it also requires the need to use technology and communicate more effectively. Juan et al. (2011) also pointed out that the online instructor becomes a specialist to guide students' learning process. In this role, instructors facilitate student learning, rather than teach students (lead lecture).

In a learner-driven classroom, it is important for the instructor to accommodate the differences of students' previous knowledge and experiences in their learning processes. This study will use word embedding and basic recurrent neural network to visualize terminological schematic context, and predict terms for translation students at different levels in their learning. In natural language processing, word embedding, or word vectorization, is a methodology to map words from human, readable vocabulary to a corresponding vector of numbers to represent word similarities and semantics, which are embedded in a web of associations. By asking learners with different cognitive backgrounds to annotate terms in a United Nations (UN) corpus, instructors can capture their specific patterns to classify words in this corpus, as well as the internal relationship, which then help them predict difficult or specialized terms for translation students.

The research questions are:

1. Can relevant corpus help translation students better leverage machine perception about word meaning in a specialized text?
2. Can machine learning (ML) methodology capture individual learners' different characteristics when they approach terminology?

Literature Review

Terminological Schematic Context

Learners can integrate new information into their existing knowledge structure in two ways: (1) top-down, starting from known features (e.g., grammatical features) inferenced from sampled data (e.g., a particular set of natural texts, or a corpus, that represents the use of language features (see Biber et al., 1998)), in which the shared, collective knowledge is passed down from instructors to each individual learner; and (2) bottom-up, with each individual learner working on sampled data, so as to extract features on their own, in which each learning process is unique from the start, as no one has the same cognitive background. In both scenarios, it is important for an instructor to incorporate learners' individual cognitive features into their learning processes.

In our research, we aim to capture translation students' cognitive capabilities when they learn terms and their associated concepts. There are two main features in translation terminology: (1) specialization, as each term represents specialized knowledge (see Faber & Leon-Arous, 2016); (2) cognitive difficulty level, as Muegge (2020) stated, "...not only the special words that belong to a specific discipline should be managed as part of every translation project, but every 'difficult' word." Among these two features, identifying difficult words for learners with multiple levels of previous knowledge and experience is usually a daunting task for an instructor, as the meaning of "difficult" varies from person to person.

Thus, we propose the concept Terminological Schematic Context (TSC), which is a subjective structure encompassing cognitive processes that help translators access relevant knowledge in their working memory by leveraging both terms and organizational structures at a higher level. Here, working memory "refers to a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks, as language comprehension, learning, and reasoning" (Baddeley 1986), and schema refers to the higher-level cognitive representation in the hierarchical structure that is apt to specify the relationships between its components at a lower level (Mazzone 2015). This schematic organization of memory allows the activation of any items spread to schemata that are the most accessible, due to previous experience. The activation of a schema, in turn, activates its other components, so as to predict a likely cognitive context for the original item.

Human Learning Versus Machine Learning

ML methodology can serve as a starting point for us to visualize TSC and predict learners' term recognition based on their previous schemata. Machine learning is, in essence, a type of statistical learning, referring to a vast set of tools for understanding data (James et al., 2013). ML shares common ground with human learning, as Dehaene (2020) phrased it: "To learn is to form an internal model of the external world....Learning is adjusting the parameters of a mental model." This definition highlights the dynamic features of a learning process, reflecting the diversity and individuality of a learner, who could be both a human being and a machine-learning system.

In our study, we will focus on one particular internal model, that is, the language model of human and machine learning systems. We will apply word embedding and basic recurrent neural network methodologies in natural language processing (NLP). NLP is a branch of artificial intelligence that helps computers understand, analyze, manipulate, and potentially generate human language (Yannakakis, et al., 2018). Word embedding converts texts to real-valued vectors that encode the meaning of the word, by looking into the context of the word. Those words that are closer in the vector space are expected to be similar in meaning.

Methods

For word embeddings, we used two datasets. One is pre-trained word vectors using Gensim, an open-source Python library for NLP, with a focus on topic modeling. We chose the model of 100-dimensional vectors, which is trained on Wikipedia data with 6 billion tokens and a 400,000-word vocabulary. The other dataset is the English version of a resolution adopted by the General Assembly on 16 April 2021 (Symbol A/RES/75/271) with 2,266 words in total. We used the Word2Vec model (see Mikolov 2013a, 2013b) to train the UN document and also visualize the vectors.

Then we used the UN document to train our term recognition models for different levels of learners. Our training data is divided into three categories based on learners' existing linguistic and cultural competence, namely, beginner learners (BL), intermediate learners (IL), and advanced learners (AL). Lacking the competence of identifying phrases that have specialized meaning in this context, beginner learners usually focus on individual words, for example, "biodiversity," rather than "biodiversity conservation. Advanced learners usually focus on specialized phrases, rather than individual words, as these words are already in their memory. Intermediate learners are somewhere in between. For example, as seen in Table 1, they are able to identify "General Assembly" as a specialized phrase, but not "biodiversity conservation. In the annotation process, we referred to TextRazor, a NPL meaning extraction tool, to get some insights. We used the "sequential" model as our framework for the term recognition model training. The programming language is Python.

Table 1

Learners' Annotation of the Terms

BL	BL TERM	IL	IL TERM	AL	AL TERM	en_US
0	none	0	none	0	none	nature knows no borders: transboundary cooperation -
1	biodiversity	0	none	1	biodiversity conservation	a key factor for biodiversity conservation
1	restoration	0	none	0	none	restoration and sustainable use
0	none	0	none	0	none	United Nations
1	assembly	1	general assembly	0	none	General Assembly
0	none	0	none	0	none	seventy-fifth session

Results

Regarding the first research question, we find that words learned from a more general corpus (Wikipedia data) are distributed differently in the vector space, compared with those learned from a more specialized corpus (the UN document). Specialized corpus can provide translation students with more specific TSC. For example, when we ask the program to return the most similar word of the term “biodiversity,” we find that Wiki embeddings return a different set of words (see Figure 1), compared with what the UN Corpus embeddings return (see Figure 2). However, we can see that the similarity in the UN text scenario is not as high as that in the Wikipedia one (e.g., 0.3023089 vs. 0.7893801). This might be due to the fact that the corpus size of the UN document is too small (2,266 words), so that the system has very limited data to learn from.

Figure 1

Most Similar Words for “Biodiversity” Learned from Wikipedia

```
[('ecosystems', 0.78935801
 ('ecosystem', 0.762296438
 ('conservation', 0.724524
 ('ecological', 0.72449582
 ('habitat', 0.70174437761
 ('ecology', 0.69636100530
 ('wetlands', 0.6960740685
 ('wildlife', 0.6822534203
 ('wetland', 0.68006634712
 ('fauna', 0.6731040477752
```

Most Similar Words for “Biodiversit … Document

```
 ('people', 0.24367001652
 ('importance', 0.231840
 ('ecosystems', 0.224160
 ('existing', 0.22342425
 ('plans', 0.22093452513
 ('areas', 0.21552571654
 ('initiatives', 0.19993
```

We can also use the word embeddings to do math, as texts have been converted to numbers. For example, “biodiversity + cooperation” equals “environment” in the wiki embeddings, whereas “initiatives” in embeddings are learned from the UN document. Again, even though the probability is much higher in the Wikipedia case, the results indicate that the more specialized corpus will teach the word embedding model that is more specific to TSC.

Figure 3

Biodiversity + Cooperation = Environment in Wiki Embeddings

```
[('environment', 0.7203084230422974), ('development', 0.7139549255371094), ('conservation',
0.7094401717185974), ('sustainable', 0.6953283548355103), ('climate', 0.6911009550094604)]
```

Figure 4

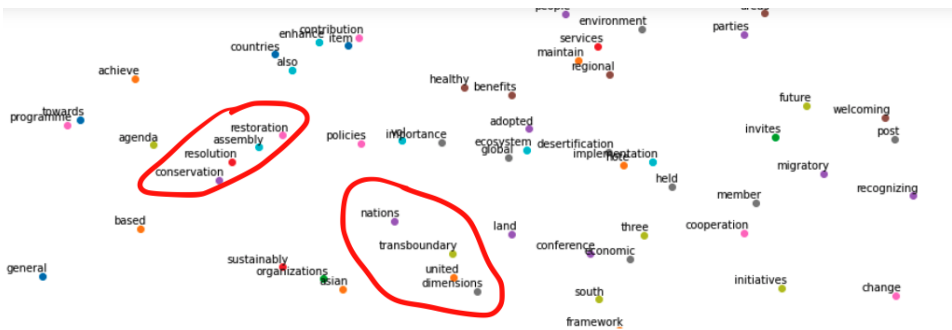
Biodiversity + Cooperation = Initiatives in Word Embeddings Learned from the UN Document

```
[('initiatives', 0.3535212278366089), ('strategic', 0.2952273190021515), ('people', 0.28827
51524448395), ('importance', 0.28005051612854004), ('stakeholders', 0.2460464984178543)]
```

There are some interesting machine perceptions that we can leverage. For example, “restoration,” “assembly,” “resolution,” and “conservation” are clustered together in our UN document-trained word vector space (see figure 5), which makes sense to some degree, if you want to establish the logical relationship between these words in an intuitive way. For example, this resolution is held in an assembly, and its topic is conservation and restoration of the environment.

Figure 5

Part of the Visualization of Learned Words based on the UN Document



Regarding the second research question, we used the basic RNN to build a model (see Figure 6) and find that the precision/recall/accuracy scores are reasonable, in terms of predicting terms identified by different learners (see Figure 7).

Figure 6

Basic RNN Model Framework

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, None, 32)          14528
_________________________________________________________________
lstm (LSTM)                  (None, 32)                8320
_________________________________________________________________
dense (Dense)                (None, 32)                1056
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 33
=================================================================
Total params: 23,937
Trainable params: 23,937
Non-trainable params: 0
```

Figure 7
Term recognition prediction for beginner learners

```
7/7 [==============================] - 0s 26ms/step - loss: 0.6192 - accuracy: 0.5843 -
precision_m: 0.5765 - recall_m: 1.0000 - val_loss: 0.6155 - val_accuracy: 0.6607 - val_p
recision_m: 0.6615 - val_recall_m: 1.0000
Epoch 8/10
7/7 [==============================] - 0s 29ms/step - loss: 0.5427 - accuracy: 0.6977 -
precision_m: 0.6786 - recall_m: 1.0000 - val_loss: 0.5984 - val_accuracy: 0.6429 - val_p
recision_m: 0.6542 - val_recall_m: 0.9688
Epoch 9/10
7/7 [==============================] - 0s 25ms/step - loss: 0.5143 - accuracy: 0.7786 -
precision_m: 0.7244 - recall_m: 0.9926 - val_loss: 0.5816 - val_accuracy: 0.6429 - val_p
recision_m: 0.6776 - val_recall_m: 0.8899
Epoch 10/10
7/7 [==============================] - 0s 28ms/step - loss: 0.4376 - accuracy: 0.8304 -
precision_m: 0.7798 - recall_m: 0.9879 - val_loss: 0.5688 - val_accuracy: 0.6786 - val_p
recision_m: 0.7077 - val_recall_m: 0.8899
```

Discussion and Conclusion

This study shows the application of ML methods in a human learning process, which helps instructors to predict translation students' cognitive status for terms in a specialized corpus. It also helps the students visualize their TSC. While this study is focused on term/concept comprehension for translation students, it can be applied to other languages in a language teaching classroom. As we mentioned before, we annotated the data based on perceptions provided by an NLP application. We could involve real-life learners to annotate the data in our follow-up studies.

As Webb et al. (2020) pointed out, ML systems are infiltrating our lives and are beginning to become important in our educational systems. While we are beginning to use ML methodology and framework to innovate our pedagogy, like what this study has demonstrated, we should never ignore the role of instructors in a human-learning process. Now, it is never more appropriate to ask the question: What is the true value of an instructor under the influence of ML? Nowadays, in a world where everything is digitalized, and much easier for people to access, providing structured information or knowledge is no longer the fundamental reason why the role of a teacher should exist. I argue that the true value of an instructor lies in the communicative and the human part of teaching. Human instruction is helpful, and needed, not just because of the quality and quantity of information it can provide, but because of the empathy instructors share with their learners for them to communicate with each other. Humans share similar experiences or feelings about language as well, as in the physical and mental worlds it is associated with. Starting from this common ground, learners can verify their hypotheses or inference models from an advanced learner, so that they can generalize them on a larger scale, or transfer these models to different tasks. ML technologies help instructors to focus on the human part of teaching, rather than replacing humans' instruction. They are useful tools for instructors to understand multiple levels of learners, so as to maximize the empathetic areas with the learners.

References

Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511804489

Choi, H. J., & Park, J. (2006). Difficulties that a novice online instructor faced: A case study. *The Quarterly Review of Distance Education*,7, 317–322.

Faber, Pamela and León-Araúz Pilar. 2016. "Specialized Knowledge Representation and the Parameterization of Context." *Frontiers in Psychology*. 2016 (7), 196.

Fein, A. D., & Logan, M. C. (2003). Preparing instructors for online instruction. New Directions for Adult and Continuing Education, 100, 45–55.

Kebritchi, M., Lipschuetz, A., & Santiague, L. (2017). Issues and Challenges for Teaching Successful Online Courses in Higher Education. *Journal of Educational Technology Systems*, *46*(1), 4–29. https://doi.org/10.1177/0047239516661713

James, G., Witten, D., Hastie, T, & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer.

Juan, A. A., Steegman, C., Huertas, A., Martinez, M. J., & Simosa, J. (2011). Teaching mathematics online in the European area of higher education: An instructor's point of view. *International Journal of Mathematical Education in Science and Technology*, 42, 141–153.

Mazzone, M. (2015). Constructing the context through goals and schemata: top-down processes in comprehension and beyond. *Front. Psychol.* 6 (651): 64-76.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013a). Efficient Estimation of Word Representations in Vector Space. CoRR, abs/1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013b). Distributed Representations of Words and Phrases and their Compositionality. In NIPS (pp. 3111--3119) . Curran Associates, Inc.

Muegge, U. (2020). Why you should care about terminology management - even if you never translate a single term. *ATA Chronicle, 49*(2), 22-25.

TextRazor Ltd. https://www.textrazor.com/

United Nations Official Document System. https://documents.un.org/prod/ods.nsf/home.xsp

Webb, ME, Fluck, A , Magenheim, J, Malyn-Smith, J, Waters, J, Deschenes, M & Zagami, J. (2020). Machine learning for human learners: opportunities, issues, tensions and threats. *Educational Technology Research and Development*. 1-22 , https://doi.org/10.1007/s11423-020-09858-

Yannakakis, G. N., & Togelius, J. (2018). *Artificial intelligence and games.* Springer International Publishing. https://doi.org/10.1007/978-3-319-63519-4