Geometric models reveal the hidden structure of conceptual knowledge

Paxton C. Fitzpatrick and Jeremy R. Manning*

Dartmouth College

*Corresponding author: jeremy.r.manning@dartmouth.edu

Abstract

We develop a mathematical framework, based on natural language processing models, for tracking and characterizing the acquisition of conceptual knowledge. Our approach embeds each concept in a high dimensional representation space, where nearby coordinates reflect similar or related concepts. We tested our approach using behavioral data collected from a group of college students. In the experiment, we asked the participants to answer sets of quiz questions interleaved between watching two course videos from the Khan Academy platform. We applied our framework to the videos' transcripts, and to text of the quiz questions, to quantify the content of each moment of video and each quiz question. We used these embeddings, along with participants' quiz responses, to track how the learners' knowledge changed after watching each video. Our findings show how a limited set of quiz questions may be used to construct rich and meaningful representations of what each learner knows, and how their knowledge changes over time as they learn.

Keywords: education, learning, knowledge, concepts, natural language processing

Introduction

How do we acquire conceptual knowledge? Memorizing course lectures or textbook chapters by rote can lead to the superficial *appearance* of understanding the underlying content, but achieving true conceptual understanding seems to require something deeper and richer. Does conceptual understanding entail connecting newly acquired information to the scaffolding of one's existing knowledge or experience [1, 5, 7, 8, 19]? Or weaving a lecture's atomic elements (e.g., its component words) into a structured network that describes how those individual elements are related? Conceptual understanding could also involve building a mental model that transcends the meanings of those individual atomic elements by reflecting the deeper meaning underlying the gestalt whole [12, 14, 18].

The difference between "understanding" and "memorizing," as framed by the researchers in education, cognitive psychology, and cognitive neuroscience [9–11, 14, 18] has profound analogs in the fields of natural language processing and natural language understanding. For example, considering the raw contents of a document (e.g., its constituent symbols, letters, and words) might provide some information about what the document is about, just as memorizing a passage might be used to answer simple questions about the passage [e.g., whether it might contain words related to furniture versus physics; 2, 3, 13]. However, modern natural language processing models [e.g., 4, 6, 17] also attempt to capture the deeper meaning *underlying* those atomic elements. These models consider not only the co-occurrences of those elements within and across documents, but also patterns in how those elements appear across different scales (e.g., sentences, paragraphs, chapters, etc.), the temporal and grammatical properties of the elements, and other high-level characteristics of how they are used [15, 16]. According to these models, the deep conceptual meaning of a document may be captured by a feature vector in a high-dimensional representation space, where nearby vectors reflect conceptually related documents. A model that succeeds at capturing an analog of "understanding" is able to assign nearby feature vectors to two conceptually related documents, *even when the words contained in those documents have very little overlap*.

What form might the representation of the sum total of a person's knowledge take? First, we might require a means of systematically describing or representing the nearly infinite set of possible things a person could know. Second, we might want to account for potential associations between different concepts. For example, the concepts of "fish" and "water" might be associated in the sense that fish live in water. Third, knowledge may have a critical dependency structure, such that knowing about a particular concept might require first knowing about a set of other concepts. For example, understanding the concept of a fish swimming in water first requires understanding what fish and water *are*. Fourth, as we learn, our "current state of knowledge" should change accordingly. Learning new concepts should both update our characterizations of "what is known" and should also unlock any now-satisfied dependencies of that newly learned concept so that they are "tagged" as available for future learning.

Here we develop a framework for modelling how knowledge is acquired during learning. The central idea is to use text embedding models to define the coordinate systems of two maps: (a) a *knowledge map* that describes the extent to which each concept is currently known and (b) a *learning map* that describes the extent to which each concept could be learned. Each location on these maps represents a single concept, and the geometries are defined such that related concepts are located nearby in space. We use this framework to analyzing and interpreting behavioral data collected from an experiment that has participants watch and answer conceptual questions about a series of recorded course lectures.

Our primary research goal is to advance our understanding of what it means to acquire deep real-world conceptual knowledge. Traditional laboratory approaches to studying learning and memory (e.g., list learning studies) often draw little distinction between memorization and understanding. Instead, these studies typically focus on whether information is effectively encoded or retrieved, rather than whether the information is *understood*. Approaches to studying conceptual learning, such as category learning experiments, can start to investigate the distinction between memorization and understanding, often by training participants to distinguish arbitrary or random features in otherwise meaningless categorized stimuli. However the objective of real-world training, or learning from life experiences more generally, is often to develop new knowledge that may be applied in *useful* ways in the future. In this sense, the gap between modern learning theories and modern pedagogical approaches and classroom learning strategies is enormous: most of our theories about *how* people learn are inspired by experimental paradigms and models that have only peripheral relevance to the kinds of learning that students and teachers actually seek. To help bridge this gap, our study uses course materials from real online courses to inform, fit, and test models of real-world conceptual learning.

Results

Discussion

Materials and methods

Participants

Experiment

Analysis

Constructing text embeddings of multiple videos and questions

Estimating held-out conceptual knowledge

Creating knowledge and learning map visualizations

References

[1] Blaye, A., Bernard-Peyron, V., Paour, J.-L., and Bonthoux, F. (2006). Category flexibility in children: distinguishing response flexibility from conceptual flexibility; the protracted development of taxonomic representations. *European Journal of Developmental Psychology*, 3(2):163–188.

[2] Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the International*

Conference on Machine Learning, pages 113–120, New York, NY. Association for Computing Machinery.

[3] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

[4] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. *arXiv*, 2005.14165.

[5] Caramazza, A. and Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, 7(8):354–361.

[6] Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., and Kurzweil, R. (2018). Universal sentence encoder. *arXiv*, 1803.11175.

[7] Constantinescu, A. O., O'Reilly, J. X., and Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292):1464–1468.

[8] Deacon, D., Grose-Fifer, J., Yang, C. M., Stanick, V., Hewitt, S., and Dynowska, A. (2004). Evidence for a new conceptualization of semantic representation in the left and right cerebral hemispheres. *Cortex*, 40(3):467–478.

[9] Gallagher, J. J. (2000). Teaching for understanding and application of science knowledge. *School Science and Mathematics*, 100(6):310–318.

[10] Hall, R. and Greeno, J. (2008). *21st century education: A reference handbook*, chapter Conceptual learning, pages 212–221. Sage Publications.

[11] Katona, G. (1940). *Organizing and memorizing: studies in the psychology of learning and teaching*. Columbia University Press.

[12] Kintsch (1970). *Learning, memory, and conceptual processes*. Wiley.

[13] Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240.

[14] Maclellan, E. (2005). Conceptual learning: the priority for higher education. *British Journal of Educational Studies*, 53(2):129–147.

[15] Manning, J. R. (2020). Context reinstatement. In Kahana, M. J. and Wagner, A. D., editors, *Handbook of Human Memory*. Oxford University Press.

[16] Manning, J. R. (2021). Episodic memory: mental time travel or a quantum "memory wave" function? *Psychological Review*, 128(4):711–725.

[17] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*, 1301.3781.

[18] Scott, P., Asoko, H., and Leach, J. (2007). *Handbook of research on science education*, chapter Student conceptions and conceptual learning in science. Routledge.

[19] Simon, M. A., Tzur, R., Heinz, K., and Kinzel, M. (2004). Explicating a mechanism for conceptual learning: elaborating the construct of reflective abstraction. *Journal for Research in Mathematics Education*, 35(5):305–329.