Searching for Argument-Counterargument Relationships in Vector Embeddings

Cherrie Chang and Joshua R. de Leeuw
Cognitive Science Department, Vassar College, Poughkeepsie, NY, USA.
cchang@vassar.edu, jdeleeuw@vassar.edu

Abstract

Vector embedding spaces are representational structures that can capture both the similarity relationship between items and various other semantic relationships. Current state-of-the-art embedding models can generate embedding vectors for individual words and longer strings of text, enabling vector spaces to encode the similarity between entire documents of text. We investigated three embedding models to see if semantic relationships besides similarity are represented in these spaces across three embedding models, focusing on the relationship between arguments and counterarguments as a specific example. While there was not a linear subspace that captured the semantic relationship between an argument and its counterargument, we found that neural networks with a single hidden layer could partially learn the transformations between an argument's embedding and the corresponding counterargument's embedding in all three spaces. The trained models generalized across three different datasets of arguments, suggesting these partially learned transformations are applicable to arguments and counterarguments in general, not just tied to the semantic context of the models' training dataset. This approach has practical applications in designing information retrieval systems for intelligent agents and, potentially, in models of cognition that use vector embedding spaces as a representational structure.

Keywords: semantic search; word embeddings; vector spaces

Introduction

Word embeddings are a powerful representational structure for semantics. These spaces typically represent words or part-words, together called tokens, as high-dimensional vectors. Constructing the space using methods like word co-occurrence (Globerson et al., 2004) and/or neural networks trained to model each word in a paragraph as a product of the conditional probabilities of the words preceding it (Bengio et al., 2000) results in desirable representational properties, such as vectors that are similar to one another having similar semantic meaning. These vector-based representations also show up in cognitive models of memory (Jones et al., 2006; Kelly & West, 2012; Kleyko et al., 2023), categorization (Surkova et al., 2020), and semantics (Grand et al., 2022).

One remarkable aspect of these spaces is word embedding arithmetic (Mikolov et al., 2013; Pennington et al., 2014), in which it was shown that simple linear combinations of vectors, like addition and subtraction, results in semantically sensible changes. For instance, starting with the vector for "Madrid", subtracting *vector*("Spain") from it, then adding *vector*("France") to it results in a vector very similar to the *vector*("Paris") (Ngo et al., 2016; Palangi et al., 2016; Parwita & Siahaan, 2019). This shows embedding spaces represent relationships between tokens in a systematic way.

Until relatively recently, most embedding spaces focused on representing individual tokens (words or part-words) in the space. With this approach, the semantics of an entire sentence, paragraph, or chapter of a book would be extracted by a model operating on the individual token embeddings (e.g. Wang et al., 2018). Recent approaches to constructing embeddings have introduced methods for generating embeddings for strings of multiple tokens. For example, the text-embedding-ada-003 model has a maximum window length of 8,192 tokens (Zhuang, 2024), allowing it to generate an embedding that can represent a large chunk of a document of text.

The possibility of embedding spaces that can capture semantic concepts expressible only by combining individual tokens into sentences or paragraphs raises questions about the representational structure of these spaces. Do these spaces capture semantic relationships that are not reducible to the relation between individual tokens? If so, are there linear subspaces that do it? Or are there non-linear transformations we can use to discover these relationships?

We set out to explore these ideas for a particular kind of complex semantic relationship: the relationship between arguments and counterarguments. We were interested in this relationship for a few reasons. First, it is certainly a relationship that is not reducible to one or even a few tokens — to make an argument, you would need a sentence at the very least; Second, the relationship is multifaceted, with many kinds of potential arguments and counterarguments for any given topic; Third, and finally, there are interesting applications for an embedding space that can capture argument-counterargument relationships in the field of argument mining (L. Li et al., 2017; Reimers et al., 2019). For example, a popular technique for giving large language models contextually relevant information that is not part of their training data is vector-based document retrieval (Caid et al., 1995; Roy et al., 2016). This typically involves storing a set of documents indexed by their embeddings in some embedding space. The language model can then be prompted, e.g. through a chat dialog, and the embedding of the prompt can be compared to all of the documents. The most similar documents to the prompt embedding may contain relevant information, so the text of those documents can be passed to the LLM as part of a modified prompt.

While this is a useful and widely adopted method, similarity is only one kind of relation that we might want to search with. In the case of an agent that has the goal to provide counterarguments to the statements that it receives, the most similar documents might not be the most useful, as

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Table 1: Example argument-counterargument pair of rows in the ArguAna dataset

	statement	*pair_id*	*type*	*stance*	*topic*	*category*
0	The minimum wage aids in the propagation of social justice and the fair treatment of workers. Businesses operating in a free market are…	0	point	PRO	business-economic …-minimum-wage	economy
1	There is no social justice in denying people the ability to work. The minimum wage serves…	0	counter	CON	business-economic-...-minimum-wage	economy

they are likely to express views that align with the prompt rather than oppose it. Instead, we want to find documents that fulfill the particular kind(s) of semantic relation(s) that exist between arguments and counterarguments.

In this paper, we test this idea in three studies, using three different datasets of natural language arguments and three different embedding spaces. In Study 1 we attempt to find a linear subspace that captures the argument-counterargument relation using PCA. In Study 2, we use simple neural network models to find a non-linear relationship between arguments and counterarguments in the embedding spaces. Finally, in Study 3 we explore the ability of the trained models to generalize to different datasets.

Datasets

We used three datasets of arguments and counterarguments.

The ArguAna Counterargs Corpus

For Study 1 and Study 2, we used the training dataset from the ArguAna Counterargs Corpus, an English corpus constructed by the NLP Group at Leibniz University Hannover, Bauhaus-Universität Weimar and affiliations to study the retrieval of counterarguments given an argument (*About – ArguAna*, n.d.). The corpus has been used in similar literature in information retrieval (e.g. Hashemi et al., 2023), and is an important dataset used in many machine learning benchmarks (Thakur et al., 2021). It was chosen due to its large size and representativeness of real-life debates: it contains 6,753 argument-counterargument pairs crawled from idebate.org, divided into 15 broad debate categories, such as “Science” and “Culture”. In each category are ~30 debates (Wachsmuth et al., 2018), with each debate comprising a labeled list of arguments ~100 words long labeled for or against the debate’s thesis statement. After every argument is a counterargument that corresponds to it.

Each debate contains 7 argument-counterargument pairs on average [min=3, max=20]. The dataset was then converted into a table with each row representing one argument. An example row in the table would include a “statement” column for the argument or counterargument text, a “debate” column identifying the debate topic it belongs to, a “pair_id" unique to the argument and its counterpart in the debate, a “type” column for whether the row represents an argument or a counterargument, and a “stance” column determining whether the argument is for or against the debate’s thesis statement (Table 1).

The IBM GPR-KB-55 Dataset

For Study 3, we used one of the IBM Project Debater’s datasets, the GPR-KB-55, which is composed of 55 arguments and their corresponding counterarguments written by an expert human debater (Orbach et al., 2019). These arguments are “general-purpose”, in that they are topic-agnostic and broadly capture the general structure of an argument and its counterargument without holding any topic-specific information. An example argument from this dataset would be “We need to think about how this affects us right now;” and its corresponding counterargument would be “The long-term effects in this case greatly outweigh the short-term ones” (Table 2). This dataset was chosen as an alternative from the ArguAna dataset to test the models on, since it follows a similar paired argument-counterargument structure. The topic-agnostic nature of the arguments may also serve as a baseline for determining whether the models have learned a relation generalizable across all topics.

Table 2: Example rows in the GPR-KB-55 dataset

	argument	*counterargument*
0	We need to think about how this affects us right now.	The long-term effects in this case greatly outweigh the short-term ones.
1	<ACTION> <TOPIC> will benefit us in the future.	There are many things that could theoretically benefit us in the future. Unfortunately we have to…

The IBM EACL Dataset

The EACL dataset is another dataset from IBM’s Project Debater we used to test the models in Study 3. This dataset is composed of 2,394 labeled arguments across 55 topics manually extracted from Wikipedia (Bar-Haim et al., 2017). The arguments in the dataset are manually annotated as for or against their debate topic’s thesis, but are not paired. For example, the thesis “This house would introduce year round schooling” contains 11 arguments for it (e.g. “Parents are in favor of the year-round schedule”) and 5 counterarguments (e.g. “If schools are open longer the operating and maintenance costs may increase”) against it. On average, each topic has 24 arguments [min=2, max=121] and 19 counterarguments [min=2, max=161]. The total number of arguments in each topic also vary, with each topic consisting of ~44 [min=4, max=182] arguments and counterarguments.

This structure lets us test the models in a more permissive way, analyzing if given the embedding of a particular argument, a model is able to generate an embedding vector specifically near or within the subspace of counterarguments in its topic, as opposed to any vector within the broader subspace of the topic itself.

Table 3: Example rows in the EACL dataset

	topicID	*stance*	*argument*
0	644	PRO	Parents are in favor of the…
1	644	CON	If schools are open longer….

Embeddings

The embedding spaces generated by 3 embedding models, text-embedding-ada-002 (ada-002), text-embedding-3-small (ada-003-small) and Nomic Embed, were investigated. Both ada- models are from OpenAI, the company behind ChatGPT. ada-003-small succeeds ada-002 and is the "lite version" of OpenAI's newest, most powerful embedding model. Nomic Embed is from Nomic AI, a start-up focused on building explainable and accessible AI, positioned to be as powerful as ada-003 based on popular benchmark scores like MTEB (Greene et al., 2022; Nussbaum et al., 2024; Zhuang, 2024). All three models are English-based and general-purpose. ada-002 has a token window of 2048 tokens (roughly 2-3 pages of English); while ada-003-small and Nomic Embed both have a token window 4x as large at 8192 tokens. Both ada- models generate embeddings with a max dimensionality of 1536; while Nomic Embed generates up to 768-dimensional embeddings. The most invaluable thing about Nomic Embed is that it is fully reproducible, open-source, open-weights and open-data; unlike the closed source ada- models. Using Nomic Embed means our results can be further investigated by looking into the model's source code, weights and training datasets; while using and comparing the state-of-the-art, widely adopted ada- models (X. Li et al., 2024; Patil et al., 2023) give us a sense of how fast current capabilities of embedding spaces are growing.

Each of these embedding models takes in a string of text and turns it into an embedding vector, which positions the text in the model's high-dimensional embedding space. The closer the distance between two embeddings in this space, the more similar their semantics are. We decided to use cosine similarity as our similarity metric for measuring this distance, due to its wide adoption in embedding space semantic search (e.g. Thongtan & Phienthrakul, 2019), including seminal papers such as Bolukbasi et al. (2016) and Mikolov et al. (2013) where the systematic linear relationship between word vectors with related semantics was first reported. It is also advised to choose the same similarity metric used to train the models (Schwaber-Cohen, 2023; Sitikhu et al., 2019), and cosine similarity was used as a training metric in all three embedding models (Greene et al., 2022; Nussbaum et al., 2024; Zhuang, 2024).

Study 1: Linear Subspaces

We started with analyzing whether the embedding spaces already capture the semantic relation between arguments and counterarguments through a linear subspace. To test this, we ran principal components analysis (PCA) on a dataset of argument-counterargument pairs and examined whether the first few components could account for a substantial portion of the variance. Our approach is motivated by the historical success of using PCA to find relations between words and concepts. Examples include the visibly linear relationships in two-dimensional PCA projections of country vectors and capital vectors (see Mikolov et al., 2013); and the gender-coding of professions, where roles like "homemaker" and "hairdresser" are projected onto the she-gender direction in the embedding space, while "captain" and "boss" share the he-gender direction in their projections, exposing a gender bias in the embedding space captured as linear transformations to these professions from gender words (see Bolukbasi et al., 2016).

Since it is evident that PCA projections uncover the vector representations of these semantic relations between single word embeddings, it is reasonable to next see if PCA projections of semantically-related paragraph embeddings also uncover a linear relationship between them, which would mean current embedding spaces do represent more complex semantic relationships, such as those between arguments and counterarguments. This is especially viable in the current technological landscape, where embedding models are able to generate embeddings for larger and larger bodies of text, as demonstrated by the token window size difference between ada-002 and its successor, ada-003-small (Greene et al., 2022; Zhuang, 2024).

Method

We generated embeddings for each statement in the ArguAna dataset (described above) using all three models separately. These statements are paired into arguments and corresponding counterarguments. We took each pair, calculated the mean, then generated two opposite-direction embeddings by subtracting the mean from the argument embedding and counterargument embedding respectively (Bolukbasi et al., 2016). This is important for identifying subspaces as it forces the PCA to work with the differences along the relevant argument-counterargument characteristic, not other aspects of the embeddings (e.g. the topic).

Results

After removing statements that lacked a counterpart, we ran PCA with 10 components on the remaining three sets of ArguAna 8,130 embeddings (one from each model). If there is a linear subspace, we would expect to see the first PCA components explain substantially more variance than the others, because the mean-centering approach above would result in argument and counterargument vectors that point in opposite directions, and all lines joining arguments and their counterarguments would point in approximately the same

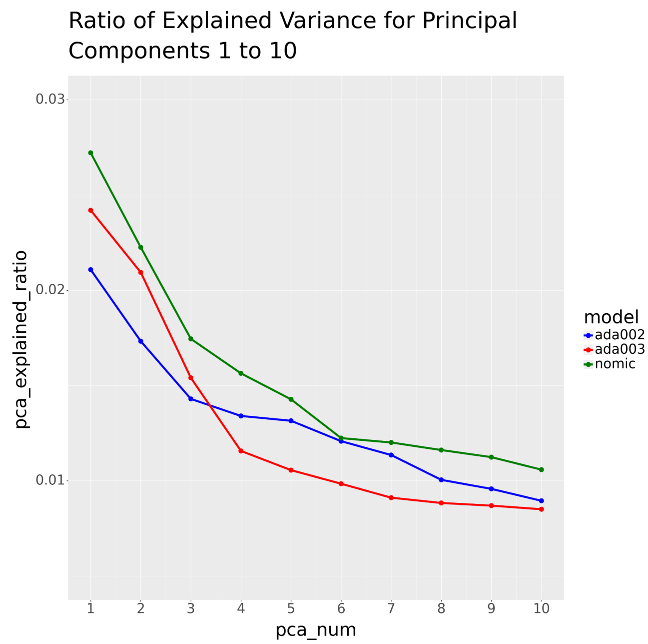

Figure 1: The ratio of explained variance explained by each of the 10 principal components we extracted, ordered from the component explaining the most variance to the one explaining the least, for all three embedding models.

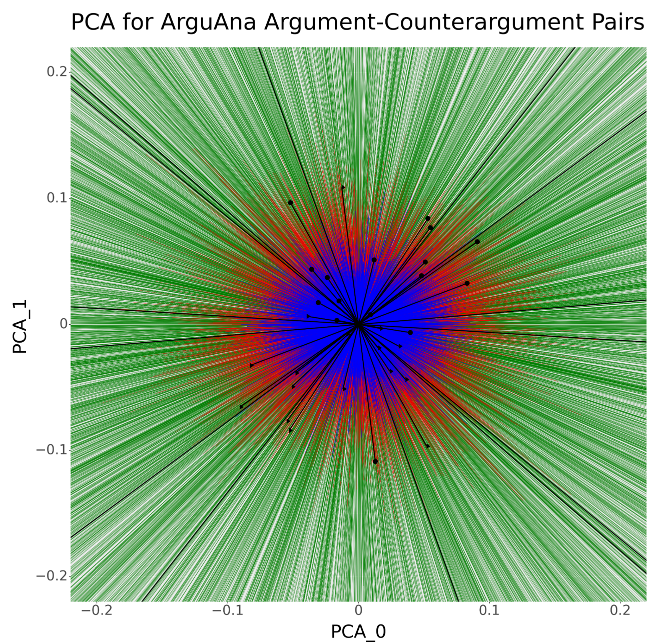

Figure 2: The projection of the argument-counterargument pairs in ArguAna onto the first two PCA components (PCA_0, PCA_1) for all three models (ada-002 in blue, ada-003 in red, nomic in green). Each pair is connected with a straight line, and pairs belonging to the "business economic policy economy general house believes national minimum wage" topic are highlighted in black. Circular markers mark the argument, and triangular markers mark the counterargument for each pair in this topic.

directions in the embedding space. Instead, the first components explained relatively little (~2.4%) variance, and differences between first and second components were low across all three models (Figure 1). Figure 2 visualizes the embedding spaces projected onto the first two components.

Discussion

Unlike previous work on embedding spaces that found linear operations to move between conceptually related words (e.g. KING-MAN+WOMAN=QUEEN), none of the embedding spaces showed clear linear subspaces capturing the relationship between arguments and counterarguments. The lines connecting arguments and counterarguments in Figure 2 visualize this, pointing in different directions even in a single topic. This may not be surprising, as arguments and counterarguments represent a complex set of relations, with many possible kinds of pairings (Bentahar et al., 2010).

Study 2: Nonlinear Transformations

In our second study, we see if there are consistent nonlinear relationships between argument and counterargument pairs.

Method

We used the same 4,074 embedding pairs from the ArguAna training set as Study 1 for training, and generated 1403 embedding pairs from the ArguAna test set to use for validation. We did not mean-center either embedding sets. Instead, we used a supervised learning approach to try and learn a mapping between an argument's embedding vector and its counterargument's embedding vector. To do this, we trained a three-layer (input, hidden, and output) neural network to take an embedding vector as input and generate as output its counterargument embedding, repeated for all 3 embedding models. All three layers had the same number of units as the embedding dimensions. Each hidden layer used a ReLU activation function. There were 4,721,664 trainable parameters in the ada- models and 1,181,184 in the Nomic model. The models were compiled with the Adam optimizer (learning rate=0.001) and cosine similarity as the loss function, then trained over 20 epochs with a batch size of 1.

The batch size of 1 was to make implementing a custom metric easier: To analyze how close a model's output embedding for an input embedding is to the target counterargument embedding, we defined a metric function that, given an output embedding from a model, calculates its cosine similarity to every embedding in the entire dataset, returning the vector closest to it. A model is considered to successfully predict a counterargument embedding if the vector it is closest to is the target embedding. This approach echoes previous work on evaluating embedding space arithmetic (e.g. Bolturk & Kahraman, 2018; Ng, 2017). We also chose this metric because it closely simulates a memory or document retrieval use case, in which a probe vector is compared to entries in a memory store to retrieve the most similar instances (Jamieson et al., 2018; Wagenpfeil et al., 2021; Zhang & He, 2019).

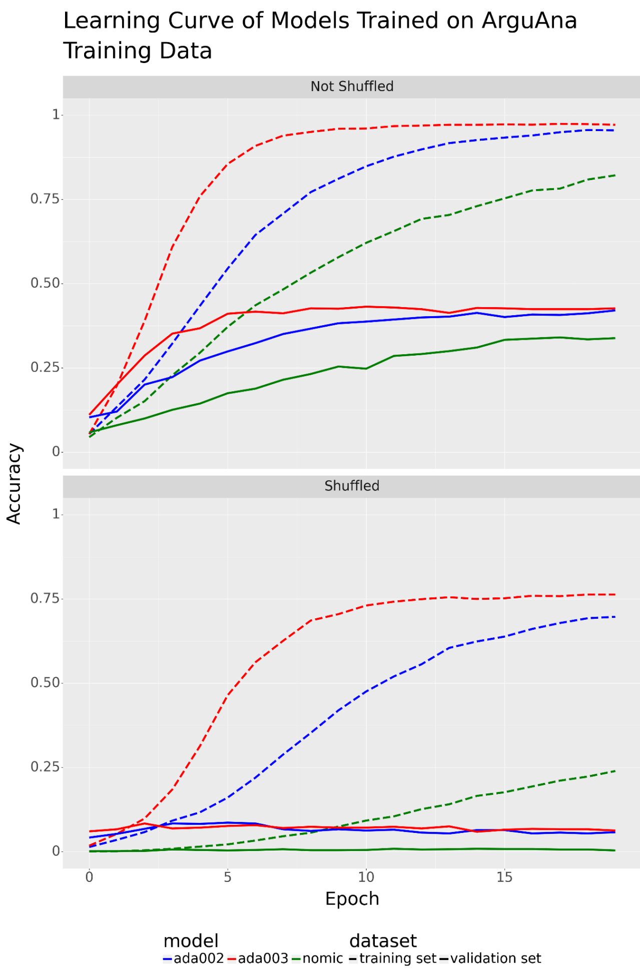

Figure 3: The accuracy of the models in predicting the correct counterargument for a given argument in the training set (dashed) and validation set (solid) over 20 epochs of training on the unshuffled (top) vs. shuffled-within-topic (bottom) versions of the ArguAna dataset respectively.

Results

All models trained on unshuffled data achieved ~39.7% [min=33.8%, max=43.2%] (Figure 3) accuracy on the held-out argument pairs after 20 epochs of training (chance performance would be 1 in 8,130). While there was clearly overfitting to the training data (~90.4% accuracy on training samples), accuracy on the validation data improved throughout training. 39.7% accuracy is perhaps too low for consistent and reliable retrieval, but it does demonstrate that even a simple 3-layer neural network architecture can partially learn the non-linear relationship between arguments and counterarguments in the embedding space.

We worried our estimate of the models' performances may be misleading because the models could learn a strategy that focused less on the relationship between arguments and counterarguments and more on preserving the topical information. For example, given an embedding vector about pro-voting-reform, the models may output a vector simply reflecting any statement about voting reform.

To test this, we trained the models again using a shuffled dataset. Within each argument topic, we randomly paired arguments with counterarguments. Thus, during training the models could only learn a general mapping between arguments and counterarguments from the same topic, but could not learn the mapping between an argument and its specific counterargument. We trained these models for 20 epochs. Performance on the hold-out validation data reached only ~5.94% [0.86%, 8.61%], well below the performance of the models trained on unshuffled data. Figure 3 shows the learning curves of all models.

Discussion

The results from our second study show that a simple three-layer neural network can partially learn a mapping between the embedding vectors for an argument and its specific counterargument, and that this performance is not unique to a particular embedding space. This also shows the models are not learning a general mapping that captures only topic-level information, but can identify specific counterarguments. Our validation data for the models included novel argument pairs *and* novel topics, suggesting that the mappings learned may be generalizable.

Study 3: Generalization

In Study 3 we explored the generalizability of the mappings learned in Study 2 to other kinds of data. If the models in Study 2 really discovered non-linear subspaces in the three embedding models' embedding spaces that capture the argument-counterargument relationship, then we would expect the models to generalize to other datasets that are more distinct from the arguments the models were trained on. To evaluate this, we tested the model on two new datasets: the GPR-KB-55 dataset and the EACL dataset. Both were constructed for IBM's Project Debater, an AI system trained to write well-structured arguments given a short description of a controversial topic (Bar-Haim et al., 2017). As described above, the GPR-KB-55 dataset contains 55 general argument-counterargument pairs that are not specific to any topic; and the EACL dataset consists of 2,394 arguments spanning 55 topics, each labeled as either an argument for the topic or a counterargument against it, but not paired with another argument or counterargument.

These two datasets therefore differ from the ArguAna dataset we trained the models on in two distinct ways: unlike the topic-tagged arguments in the ArguAna dataset, all arguments and counterarguments in the GPR-KB-55 dataset are completely general, so the models' performances on this dataset affirms whether their learned mappings between arguments and counterarguments are based on the argument-counterargument relation solely, independent from the effect of the arguments' topics; the EACL dataset provides a more permissive and realistic measure of the models' ability: the arguments and counterarguments are not paired in the dataset, so each counterargument in a topic is equally weighed as a target embedding for the models to predict from an argument embedding of the same topic. This is closer to real-world debates, where an argument can be relevantly countered by multiple counterarguments. The models' performances on the EACL dataset demonstrates whether their mappings are overfit to the training dataset's rigid one-to-one relationship between specific arguments and counterarguments, testing their robustness in fetching relevant counterarguments from more realistic documents, where arguments and counterarguments are more coarsely divided as two pools of data to draw from.

Method

We processed both datasets similarly as in Study 2, grabbing the embeddings for each statement without mean-centering them. We ran the trained models from Study 2 on every vector from both datasets, with each model predicting embeddings generated by the same embedding model as the one that generated the embeddings it was trained on. We followed the procedure in Study 2 for the GPR-KB-55 dataset, measuring the models' performance on each argument by whether their output embedding is closest to the target counterargument. For EACL, we retrieved the list of top-k embeddings from the dataset closest to the models' output embeddings for each argument. We then used the proportion of correct matches, i.e. embeddings that share the same topic as the input and are of the opposite stance, as the performance measure for the models on the input vector.

Results

Out of the 55 arguments in GPR-KB-55, the Nomic Embed model correctly output an embedding closest to the target ~8 times, giving it an accuracy of 14.5% (0.02% by chance). The ada-002 model achieved a similar accuracy level at 12.7%, but the ada-003-small model presented unexpected results, achieving only 1.81% accuracy. For EACL, the models' accuracies were measured by the proportion of correct vectors (matching topic; opposite stance) in the top-k vectors closest to its output for each argument. The results for k = 1, 2, 5, 10, 30, 200 are shown in Table 4.

The models do not seem to show consistent trends in their accuracies from k=2 to k=30, which may be explained as differences in their rate of approaching the correct vectors without precisely distinguishing similar candidate vectors.

Table 4: Top-k accuracy for argument retrieval

Proportion of retrieved arguments that are correct			
k	ada-002	ada-003-small	Nomic Embed
1	0.107	0.052	0.149
2	0.179	**0.054**	**0.156**
5	0.238	0.051	0.145
10	**0.257**	0.050	0.143
30	0.256	0.046	0.128
200	0.122	0.035	0.068

Discussion

The models did not perform as well on GPR-KB-55 as they did on the ArguAna validation set, though still better than chance (0.02%). This shows apart from ada-003-small, the models learned mappings that generalize to novel topics and, to a lesser extent, topic-agnostic arguments; suggesting the models' learned mappings partially capture the semantic relation between arguments and counterarguments, one that is able to correlate arguments and counterarguments that do not share topical context, and also delineate the argument and counterargument subspaces within a topic subspace.

General Discussion

Our three studies demonstrate that embedding spaces can partially capture semantic relationships between paragraphs beyond simple similarity. As an example, we found that while the relation between arguments and counterarguments is not captured by linear subspaces, it is discoverable via three-layer neural networks and supervised training on argument-counterargument pairs. These learned mappings between arguments and counterarguments generalized across three embedding spaces generated on three datasets, all with different argument characteristics, suggesting the mappings capture something general about this relationship.

Overall performance levels were above chance, but well short of consistent accuracy. It is unclear whether this is a limitation of the particular embedding spaces, the neural networks doing the mapping, or the dataset used to train the models. Like most machine learning problems, it is probably a little bit of all the above, and future work on this problem could explore variations on our simple model architecture and training over larger, more diverse datasets. Nevertheless, these performances are likely still strong enough to be useful in application. Vector-based document retrieval methods for providing LLMs with specific context often select several documents for any given query. A model that is 20-40% accurate at finding relevant documents would still pull useful information a lot of the time.

Beyond LLMs, we speculate that the approach of learning a non-linear transformation in embedding spaces may be fruitful in other models, like cognitive models of semantic memory. Verbal fluency test models have demonstrated that participants engage in a memory search strategy resembling switching between various patches (Lundin et al., 2023): when participants are asked to name as many food items as they can in a limited time, they may begin by naming foods sharing one kind of relationship, like "different Japanese foods", then shift to another patch of a different relation kind, like "things eaten last week". Sushi and corn may be very similar in the second reference frame, yet dissimilar in the first. This kind of change could be modeled as different transformations applied to the semantic space, such that the same probe/query results in different matches. Generalizing from this, our approach may be extended to other kinds of models that deal with the retrieval of information based on semantic relations. Whether this ends up being a useful way to think about these kinds of problems remains to be seen.

References

About – ArguAna. (n.d.). Retrieved January 31, 2024, from http://argumentation.bplaced.net/arguana/about

Bar-Haim, R., Bhattacharya, I., Dinuzzo, F., Saha, A., & Slonim, N. (2017, April 3). *Stance classification of context-dependent claims*. Conference of the European Chapter of the Association for Computational Linguistics. https://doi.org/10.18653/v1/e17-1024

Bengio, Y., Ducharme, R., & Vincent, P. (2000). A Neural Probabilistic Language Model. *Advances in Neural Information Processing Systems*, *13*. https://proceedings.neurips.cc/paper_files/paper/2000/hash/728f206c2a01bf572b5940d7d9a8fa4c-Abstract.html

Bentahar, J., Moulin, B., & Bélanger, M. (2010). A taxonomy of argumentation models used for knowledge representation. *Artificial Intelligence Review*, *33*(3), 211–259. https://doi.org/10.1007/s10462-010-9154-1

Bolturk, E., & Kahraman, C. (2018). A novel interval-valued neutrosophic AHP with cosine similarity measure. *Soft Computing*, *22*(15), 4941–4958. https://doi.org/10.1007/s00500-018-3140-y

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings* (arXiv:1607.06520). arXiv. https://doi.org/10.48550/arXiv.1607.06520

Caid, W. R., Dumais, S. T., & Gallant, S. I. (1995). Learned vector-space models for document retrieval. *Information Processing & Management*, *31*(3), 419–429. https://doi.org/10.1016/0306-4573(94)00056-9

Globerson, A., Chechik, G., Pereira, F., & Tishby, N. (2004). Euclidean Embedding of Co-Occurrence Data. *Advances in Neural Information Processing Systems*, *17*. https://proceedings.neurips.cc/paper_files/paper/2004/hash/ec1f850d934f440cfa8e4a18d2cf5463-Abstract.html

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, *6*(7), Article 7. https://doi.org/10.1038/s41562-022-01316-8

Greene, R., Sanders, T., Weng, L., & Neelakantan, A. (2022, December 15). *New and improved embedding model*. https://openai.com/blog/new-and-improved-embedding-model

Hashemi, H., Zhuang, Y., Kothur, S. S. R., Prasad, S., Meij, E., & Croft, W. B. (2023). *Dense Retrieval Adaptation using Target Domain Description* (arXiv:2307.02740). arXiv. https://doi.org/10.48550/arXiv.2307.02740

Jamieson, R. K., Avery, J. E., Johns, B. T., & Jones, M. N. (2018). An Instance Theory of Semantic Memory. *Computational Brain & Behavior*, *1*(2), 119–136. https://doi.org/10.1007/s42113-018-0008-2

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003

Kelly, M., & West, R. (2012). From Vectors to Symbols to Cognition: The Symbolic and Sub-Symbolic Aspects of Vector-Symbolic Cognitive Models. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34). https://escholarship.org/uc/item/0df6q3v5

Kleyko, D., Rachkovskij, D., Osipov, E., & Rahimi, A. (2023). A Survey on Hyperdimensional Computing aka Vector Symbolic Architectures, Part II: Applications, Cognitive Models, and Challenges. *ACM Computing Surveys*, *55*(9), 175:1-175:52. https://doi.org/10.1145/3558000

Li, L., Mao, L., & Chen, M. (2017). Word Embedding and Topic Modeling Enhanced Multiple Features for Content Linking and Argument / Sentiment Labeling in Online Forums. In G. Giannakopoulos, E. Lloret, J. M. Conroy, J. Steinberger, M. Litvak, P. Rankel, & B. Favre (Eds.), *Proceedings of the MultiLing 2017 Workshop on Summarization and Summary Evaluation Across Source Types and Genres* (pp. 32–36). Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-1005

Li, X., Henriksson, A., Duneld, M., Nouri, J., & Wu, Y. (2024). Evaluating Embeddings from Pre-Trained Language Models and Knowledge Graphs for Educational Content Recommendation. *Future Internet*, *16*(1), Article 1. https://doi.org/10.3390/fi16010012

Lundin, N. B., Brown, J. W., Johns, B. T., Jones, M. N., Purcell, J. R., Hetrick, W. P., O'Donnell, B. F., & Todd, P. M. (2023). Neural evidence of switch processes during semantic and phonetic foraging in human memory. *Proceedings of the National Academy of Sciences*, *120*(42), e2312462120. https://doi.org/10.1073/pnas.2312462120

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, *26*. https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

Ng, P. (2017). *dna2vec: Consistent vector representations of variable-length k-mers* (arXiv:1701.06279). arXiv. https://doi.org/10.48550/arXiv.1701.06279

Ngo, D. L., Yamamoto, N., Tran, V. A., Nguyen, N. G., Phan, D., Lumbanraja, F. R., Kubo, M., & Satou, K. (2016). Application of Word Embedding to

Drug Repositioning. *Journal of Biomedical Science and Engineering*, *9*(1), Article 1. https://doi.org/10.4236/jbise.2016.91002

Nussbaum, Z., Morris, J. X., Duderstadt, B., & Mulyar, A. (2024). *Nomic Embed: Training a Reproducible Long Context Text Embedder* (arXiv:2402.01613). arXiv. https://doi.org/10.48550/arXiv.2402.01613

Orbach, M., Bilu, Y., Gera, A., Kantor, Y., Dankin, L., Lavee, T., Kotlerman, L., Mirkin, S., Jacovi, M., Aharonov, R., & Slonim, N. (2019). *A Dataset of General-Purpose Rebuttal* (arXiv:1909.00393). arXiv. https://doi.org/10.48550/arXiv.1909.00393

Palangi, H., Deng, L., Shen, Y., Gao, J., He, X., Chen, J., Song, X., & Ward, R. (2016). Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(4), 694–707. https://doi.org/10.1109/TASLP.2016.2520371

Parwita, I. M. M., & Siahaan, D. (2019). Classification of Mobile Application Reviews using Word Embedding and Convolutional Neural Network. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 1–8. https://doi.org/10.24843/LKJITI.2019.v10.i01.p01

Patil, A., Han, K., & Jadon, A. (2023). *A Comparative Study of Text Embedding Models for Semantic Text Similarity in Bug Reports* (arXiv:2308.09193). arXiv. https://doi.org/10.48550/arXiv.2308.09193

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

Reimers, N., Schiller, B., Beck, T., Daxenberger, J., Stab, C., & Gurevych, I. (2019). *Classification and Clustering of Arguments with Contextualized Word Embeddings* (arXiv:1906.09821). arXiv. https://doi.org/10.48550/arXiv.1906.09821

Roy, D., Ganguly, D., Mitra, M., & Jones, G. J. F. (2016). *Representing Documents and Queries as Sets of Word Embedded Vectors for Information Retrieval* (arXiv:1606.07869). arXiv. https://doi.org/10.48550/arXiv.1606.07869

Schwaber-Cohen, R. (2023, June 30). *Vector Similarity Explained | Pinecone*. https://www.pinecone.io/learn/vector-similarity/

Sitikhu, P., Pahi, K., Thapa, P., & Shakya, S. (2019). A Comparison of Semantic Similarity Methods for Maximum Human Interpretability. *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, *1*, 1–4. https://doi.org/10.1109/AITB48515.2019.8947433

Surkova, A., Skorynin, S., & Chernobaev, I. (2020). Word embedding and cognitive linguistic models in text classification tasks. *Proceedings of the XI International Scientific Conference Communicative Strategies of the Information Society*, 1–6. https://doi.org/10.1145/3373722.3373778

Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., & Gurevych, I. (2021). *BEIR: A Heterogenous Benchmark for Zero-shot Evaluation of Information Retrieval Models* (arXiv:2104.08663). arXiv. https://doi.org/10.48550/arXiv.2104.08663

Thongtan, T., & Phienthrakul, T. (2019). Sentiment Classification Using Document Embeddings Trained with Cosine Similarity. In F. Alva-Manchego, E. Choi, & D. Khashabi (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (pp. 407–414). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-2057

Wachsmuth, H., Syed, S., & Stein, B. (2018). Retrieval of the Best Counterargument without Prior Topic Knowledge. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 241–251. https://doi.org/10.18653/v1/P18-1023

Wagenpfeil, S., Engel, F., Kevitt, P. M., & Hemmje, M. (2021). AI-Based Semantic Multimedia Indexing and Retrieval for Social Media on Smartphones. *Information*, *12*(1), Article 1. https://doi.org/10.3390/info12010043

Wang, G., Li, C., Wang, W., Zhang, Y., Shen, D., Zhang, X., Henao, R., & Carin, L. (2018). *Joint Embedding of Words and Labels for Text Classification* (arXiv:1805.04174). arXiv. https://doi.org/10.48550/arXiv.1805.04174

Zhang, M., & He, Y. (2019). GRIP: Multi-Store Capacity-Optimized High-Performance Nearest Neighbor Search for Vector Search Engine. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1673–1682. https://doi.org/10.1145/3357384.3357938

Zhuang, J. (2024, January 25). *New embedding models and API updates*. https://openai.com/blog/new-embedding-models-and-api-updates