Title
Constructing Word Meaning without Latent Representations using Spreading Activation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Shabahang, Kevin
Dennis, Simon
Yim, Hyungwook

Publication Date
2022

Peer reviewed

Constructing Word Meaning without Latent Representations using Spreading Activation

Kevin D. Shabahang (k.shabahang@gmail.com)
School of Psychological Sciences, The University of Melbourne, Australia
Hyungwook Yim (hyungwook.yim@unimelb.edu.au)
School of Psychological Sciences, The University of Melbourne, Australia
Department of Cognitive Sciences, Hanyang University, Seoul, Republic of Korea
Simon J. Dennis (simon.dennis@unimelb.edu.au)
School of Psychological Sciences, The University of Melbourne, Australia

Abstract

Models of word meaning, like the Topics model (Griffiths et al., 2007) and word2vec (Mikolov et al., 2013), condense word-by-context co-occurrence statistics to induce representations that organize words along semantically relevant dimensions (e.g., synonymy, antonymy, hyponymy etc.). However, their reliance on latent representations leaves them vulnerable to interference and makes them slow learners. We show how it is possible to construct the meaning of words online during retrieval to avoid these limitations.

We implement our spreading activation account of word meaning in an associative net, a one-layer highly recurrent network of associations, called a Dynamic-Eigen-Net, that we developed to address the limitations of earlier variants of associative nets when scaling up to deal with unstructured input domains such as natural language text. After fixing the corpus across models, we show that spreading activation using a Dynamic-Eigen-Net outperforms the Topics model and *word2vec* in several cases when predicting human free associations and word similarity ratings. We argue in favour of the Dynamic-Eigen-Net as a fast learner that is not subject to catastrophic interference, and present it as an example of delegating the induction of latent relationships to process assumptions instead of assumptions about representation.

Keywords: Retrieval; Dynamic; Associative; Semantic; Process model; Words

Landauer and Dumais (1997) echoed Plato's observation that most of our knowledge about the meaning of words depends on the induction of latent relationships between words that never directly co-occur. A statistical learning account of the formation of the meaning of words from linguistic experience must specify how to exploit surface-level co-occurrence patterns to infer relationships between words that have not appeared together. How can the system infer a relationship between two words like EAGLE and HAWK if they never co-occur, but both occur in the context of other words like FEATHER and FLY?

In the absence of direct co-occurrence, the relationship between words like EAGLE and HAWK is latent. Distributional semantics accounts of word meaning assume that latent relations directly correspond to representations in memory, and specify a set of transformations that encode the surface-level co-occurrence patterns of words into a latent representation. The latent relationships between words result from an encoding process that compresses each word's full co-occurrence history into a lower-dimensional representation.

Some accounts treat discrete documents as separate contexts, and assume that the raw input to the system is a matrix of word vectors, where each element of each word's vector is proportional to the frequency of the word in the corresponding document. Vectors for words like EAGLE and HAWK may not significantly overlap in the raw word-by-document matrix. In Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) the latent relationship between words becomes explicit when the original word vectors are projected to a lower-dimensional subspace. In the Topics model (Griffiths et al., 2007), the latent representation corresponds to a set of discrete topics, whose combination is assumed to generate each of the observed documents. The latent relationship between two words is mediated through unobserved topics to which both words correspond.

Word2vec (Mikolov et al., 2013) defines context based on a fixed-size window of text from a corpus. At each slice of text, the word in the middle of the window is designated as the target and the surrounding words are used as its context. Latent representations are formed through gradual changes to the connectivity of a multi-layer neural net that is trained to learn the conditional probability of each target word given its context. The input and output layers are local-code representations, where each unique word is assigned a unique index. The context words are activated in

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

the input layer and projected through a lower dimensional hidden layer and back out through the output layer. The lower-dimensional hidden layer acts in a similar fashion as the lower-dimensional subspace in LSA, and increases the proximity of words in the embedding based on overlap between their contexts.

Levy and Goldberg (2014) showed that transforming a word-by-context co-occurrence matrix using the shifted Pointwise Mutual Information (SPMI) and compressing the result along the context dimension approximates the same objective function used to train word2vec. As we later show, it is possible to obtain a similar level of performance, and sometimes better, compared to word2vec by using spreading activation in an associative net, instead of dimensionality reduction, for revealing latent structure. An associative net is a network of direct associations, stored in a weight matrix, coupled with a recurrence relation that specifies how activation spreads from an initially active set of nodes to the rest of the network. Spreading activation forces the structure in the network to interact with the initial activations to yield a representation that integrates those initial activations with the global associative structure.

There are several reasons why avoiding a latent representation is desirable. A system capable of inducing latent relations as a result of processing, instead of representation, is more dynamic and context-sensitive, and therefore more adaptable. Such a system is better suited for tasks like predication, where the meaning of one word is conditioned on another word, whereas models like LSA and word2vec require further augmentation. One problem with latent representations is that their formation is slow and inflexible. For example, the Topics model and word2vec both require several passes through the same input to form stable representations that capture semantic relations. As a result, it is unclear if models like word2vec and the Topics model can account for the fast-learning and adaptation to the environment characteristic of human learners (e.g., Wood et al., 2020).

Another problem is more specific to neural embedding approaches like word2vec, which rely on learning systems that are notorious for their vulnerability to interference. McClosky (1989) showed how learning one set of input-output mappings in a three-layer neural network, trained with backpropagation, completely wipes out information encoded from previously learned input-output mappings. More recently, Manning and Jones (2020) used polysemous words to show that word2vec suffers from similar problems due to interference. Polysemous words have the same spelling, but take up different meanings depending on their context. A word like BANK, has its dominant meaning as a financial institution and its secondary meaning as the land surrounding a body of water. Manning and Jones found that word2vec favoured one sense of polysemous words over another sense, depending on whether contexts portraying the one sense were trained before or after the contexts portraying the other sense. That is, if the network was trained with all the contexts using BANK to refer to a financial institution, followed by contexts that use BANK to refer to land surrounding a body of water, then the resulting word vector for BANK would be most similar to other words like RIVER or WATER instead of MONEY or ACCOUNTING. Avoiding dimensionality reduction may be one way cross-talk between encoded information can be minimized, since the distinctiveness of word representations is not lost through the compression.

In this paper, we show how latent relations can form in an associative net, using a spreading activation algorithm we developed. In prior work, we explored generalization over serial-order associations encoded from a text corpus, and found how a Dynamic-Eigen-Net was better at distinguishing congruent bigrams (e.g., “the dog”) from incongruent bigrams (“dog the”) than commonly used alternatives (Shabahang, Yim, Dennis, 2022). The Dynamic-Eigen-Net attributed greater familiarity to the congruent bigrams over incongruent bigrams, even when the weight corresponding to each of the congruent bigrams was lesioned prior to retrieval. That is, it exploited the global associative structure to generalize. Here, we extend the model to order-independent associations.

Dynamic-Eigen-Net

A Dynamic-Eigen-Net is a linear associative net with transient cue-driven weight changes. The transient weights temporarily bias the network’s settling point toward the cue and prevent runaway towards the dominant settling point of the static weight matrix. In Hebbian associative nets like the Dynamic-Eigen-Net, synapses between pairs of “neurons” strengthen when the neurons activate within a short time interval. If a single neuron encodes a single word in a sentence, then the strengthened synapses encode co-occurrence rates between words that keep the same company. Hebbian learning only requires a single exposure to the training data compared to backpropagation learning which requires many exposures before producing a stable memory representation. In addition to a memory representation for capturing associations, the weight matrix, a state vector tracks the momentary activations of the system.

Processing in an associative net is characterized by an initial state and an update function that propels the system forward in time. The state-transition law specifies how its memory weights, $\mathbf{W}$, interact with the momentary state, $\mathbf{x}^T_t$, to drive the system into the future, $\mathbf{x}^T_{t+1}$. For retrieval, an input cue, $\mathbf{x}^T_0$, is used to initialize the state for the first time-point, and the state at the next time-point is obtained as a function of the vector-matrix multiplication of the current state and the weight matrix, $\mathbf{x}^T_{t+1} = f(\mathbf{x}^T_t\mathbf{W})$. The process is carried out iteratively until further iterations have no additional effect on the state vector (i.e., when $\mathbf{x}^T_{t+1} \approx \mathbf{x}^T_t$). Retrieval forces the interconnections between words encoded from previously learned patterns to interact with the cue until the system reaches an equilibrium state. The equilibrium state is treated as the retrieved pattern.

In a Dynamic-Eigen-Net, the outer-product of the initial cue is added to the weight matrix before recurrence takes place, and is reset once the state settles to an equilibrium. In a linear associative net, each step in the recurrence pulls the state closer to a weighted combination of the eigenvectors, but ultimately settles toward the dominant eigenvector – the dimension capturing the highest amount of variance in the encoded patterns. Temporarily adding the outer-product of a cue adds another term, which persistently pulls the state towards the initial representation to prevent flight towards the dominant eigenvector. The update function for a Dynamic-Eigen-Net is given by,

$$\mathbf{x}_{t+1}^T = \frac{\mathbf{x}_t^T\left(\mathbf{W} + \mathbf{x}_0\mathbf{x}_0^T\right)}{\|\mathbf{x}_t^T\left(\mathbf{W} + \mathbf{x}_0\mathbf{x}_0^T\right)\|}$$

If we let $\mathbf{x}^{\mathrm{T}}_{\infty}$ be the state in the limit, and λ_{∞} be the primary eigenvalue of the modified weight matrix, $\mathbf{W} + \mathbf{x}_0\mathbf{x}^{\mathrm{T}}_0$, the following equation holds:

$$\mathbf{x}_{\infty}^T\left(\mathbf{W} + \mathbf{x}_0\mathbf{x}_0^T\right) = \sum_i \lambda_i\left(\mathbf{x}_{\infty}^T\dot{\mathbf{e}}_i\right)\dot{\mathbf{e}}_i^T + \left(\mathbf{x}_{\infty}^T\mathbf{x}_0\right)\mathbf{x}_0^T = \lambda_{\infty}\mathbf{x}_{\infty}^T$$

The symbol, $\dot{\mathbf{e}}_i$, denotes the *i*'th eigenvector. The right-hand side, $\lambda_{\infty}\mathbf{x}^{\mathrm{T}}_{\infty}$, follows from the fundamental eigenvalue theorem and because $\mathbf{x}^{\mathrm{T}}_{\infty}$ is the primary eigenvector of $\mathbf{W} + \mathbf{x}_0\mathbf{x}^{\mathrm{T}}_0$. The equation shows how transient weights shift the system's equilibrium, i.e., $\mathbf{x}^{\mathrm{T}}_{\infty}$, toward the initial cue. For the term, $\Sigma_i\lambda_i(\mathbf{x}^{\mathrm{T}}_{\infty}\ \dot{\mathbf{e}}_i)\dot{\mathbf{e}}_i^{\mathrm{T}}$, the eigenvectors and eigenvalues correspond to the original weight matrix, before the outer-product of the initial pattern was added. In general, the activation pattern converges towards the direction of each of the eigenvectors, weighted by its dot-product with the current state, plus the initial pattern, weighted by its dot-product to the current state. Since the states are assumed to have unit-normal length, the dot-products correspond to vector cosines (*c.f.* word embeddings).

Given a cue, the first iteration of recurrence excites words corresponding to its syntagmatic, or first-order, associates but because the resulting activations are used to probe the system again on a subsequent iteration, they in turn excite the second-order (latent) associations of the cue. Higher-order associations follow from further recurrence iterations until the system settles. Given two sentences such as "the dog played with the bone" and "the cat played with the leaf", cueing the system with CAT would activate PLAYED and LEAF in the first iteration. In the next iteration, the word PLAYED would activate DOG and BONE and so forth.

Table A1, in the appendix, shows a small corpus of fifteen different sentences that will be used to provide a toy demonstration of some of the Dynamic-Eigen-Net's properties. First each unique word in the corpus was assigned a unique index. Treating each sentence as a different context, the number of times each word co-occurred with each other word within the same context was collected into a word-by-word co-occurrence matrix, **C**, where each cell, $\mathbf{C}_{ij}$, corresponded to the number of times the *i*'th word co-occurred with the *j*'th word in the same context. In the simulations in the next section, a 7-word sliding window was used to define context.

The co-occurrence matrix is used to estimate the joint-probability of pairs of words occurring in the same context, in addition to the base-rate probability of each word. We used a smoothing parameter, *a*, when estimating the relevant probabilities. The details of the probability estimation method are shown in the appendix. We used a smoothing parameter $a = 1$ for the toy demonstration and $a = 0.4$ for the main simulations in the next section.

The Shifted Pointwise Mutual Information (PPMI) was applied to the probabilities to normalize for the different base-rate probabilities of different words (i.e., words like THE and ON occur much more often than words like DOG and CAT). For the simulations in the next section, we also remove all negative values to increase sparsity to reduce the memory load for computational reasons (*cf.* Goldberg & Levy, 2014).

The weight connecting the *i*'th word to the *j*'th word is given by,

$$\mathbf{W}_{ij} = \log_2\left(\frac{p_{ij}}{p_i p_j}\right) - \log_2(k)$$

where p_{ij} is the probability that the *i*'th word occurs with the *j*'th word and p_i and p_j are the base-rate probabilities of the *i*'th word and the *j*'th word, respectively. The ratio of the joint probability of a pair of words and the product of their base-rate probabilities ensures that the associative strength between each pair of words is proportional to the magnitude by which their joint probability exceeds their expected probability under the assumption that they occur independently.

The parameter, *k*, corresponds to the amount of "negative evidence" used to discount each association by a constant to reduce spurious associative strengths (see Levy & Goldberg, 2014). The shift parameter, *k*, was set to 1 (i.e., no negative evidence) for the toy example and 5 for the simulations in the next section.

In addition, the Dynamic-Eigen-Net requires partial inhibition of the dominant eigenvector, therefore we subtract some proportion, η, of the dominant eigenvector's outer-product, with itself, $\dot{\mathbf{e}}_{\max}\dot{\mathbf{e}}^{\mathrm{T}}_{\max}$, weighted by its corresponding eigenvalue, $\lambda_{\max}$, from the weight matrix,

$$\hat{\mathbf{W}} = \mathbf{W} - \lambda_{max}\dot{\mathbf{e}}_{max}\dot{\mathbf{e}}_{max}^T$$

The parameter, η, was set to 0 for the toy demonstration and 0.8 for the main simulations in the next section. Before cueing the system, the transient weights, $\mathbf{x}_0\mathbf{x}^{\mathrm{T}}_0$, are added to the weight matrix with a weight set to, $\lambda_{\max} + \beta\lambda_{\max}$ to ensure that the pull of the initial cue dominates during recurrence. The parameter, β, was set to 0.001. Hence the complete update function is given by,

$$\mathbf{x}_{t+1}^T = \frac{\mathbf{x}_t^T\left(\hat{\mathbf{W}} + \left(\lambda_{max} + \beta\lambda_{max}\right)\mathbf{x}_0\mathbf{x}_0^T\right)}{\left\|\mathbf{x}_t^T\left(\hat{\mathbf{W}} + \left(\lambda_{max} + \beta\lambda_{max}\right)\mathbf{x}_0\mathbf{x}_0^T\right)\right\|}$$

Overall, the model has four free parameters, a (smoothing), k (negative evidence), η (dominant eigenvector inhibition), and β (excess force of the transient weights over the dominant eigenvector).

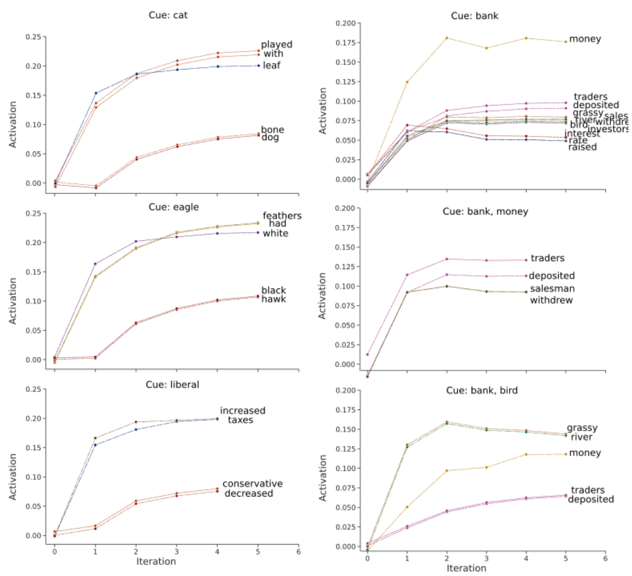

Figure 1. Trajectory of activations across iterations when the model is cued with CAT, EAGLE, and LIBERAL, shown in the left column of plots and when it is cued with BANK, BANK and MONEY, in addition to BANK and BIRD, shown in the right column of plots.

Comparing the three plots along the left column of Figure 1 with the corresponding corpus, shown in the appendix, demonstrates how latent relations between words like CAT and DOG, EAGLE and HAWK, and LIBERAL and CONSERVATIVE emerge as activation spreads through the network using the Dynamic-Eigen-Net algorithm. Only words that directly co-occurred with the cues in the same sentence are activated after the first iteration of recurrence, however, activation quickly spreads to other words that occurred in similar contexts after the second iteration. The plots along the right column show the context-sensitivity of the model. When cued with the word BANK, words with the financial sense of the cue dominate the activations since there were more cases using the word in the financial sense in the toy corpus. When further context is added, and the model is cued with BANK and MONEY, the secondary sense (i.e., river-bank) is further suppressed relative to when only the word BANK was used as the cue. Likewise, when the model is cued with both BIRD and BANK, the secondary sense dominates in the activations as shown in the lower right plot.

Simulations

To compare our account with other commonly employed models, we used the same corpus to train LSA, the Topics model, word2vec (CBOW), and the Dynamic-Eigen-Net, and examined each model's performance on the free association task, using the University of South Florida norms (USF; Nelson et al., 2004), and several word similarity datasets. Our results demonstrate how spreading activation using the Dynamic-Eigen-Net outperforms the other models in several cases and present it as a more parsimonious account. Finally, we used a set of norms, where raters categorized the relation between pairs of words into six different classes, to qualitatively profile each model in terms of their differential proclivity towards particular types of relations. Our main contribution is the demonstration that spreading activation using the Dynamic-Eigen-Net is capable of capturing the meaning of words as well as commonly used alternatives that rely on latent representations.

To construct LSA vectors, we first normalized the raw word-by-document co-occurrence matrix, $\mathbf{C}$, into the transformed matrix, $\mathbf{G}$, using,

$$\mathbf{G}_{ij} = \log_2\left(\mathbf{C}_{ij} + 1\right)\left(1 - \mathbf{H}_i\right) \text{ , where,}$$

$$\mathbf{H}_i = -\frac{1}{\log_2(D)} \sum_{j=1}^{D} \left[\frac{\mathbf{C}_{ij}}{\sum_{k=1}^{D} \mathbf{C}_{ik}} \log_2\left(\frac{\mathbf{C}_{ij}}{\sum_{k=1}^{D} \mathbf{C}_{ik}} \right) \right]$$

We then applied Singular Value Decomposition (SVD) to reduce the dimensionality of each $\mathbf{G}_i$ from the original 37,650 to 700.

To train the Topics model, we used the same procedure as Griffiths et al. (2007). We fixed the number of topics at K=1700 and set the smoothing parameters over documents and words to $50/K$ and $200/V$, respectively. For the Gibbs sampling, we used 800 burn-in samples. After the burn-ins, we used 8 samples for our estimate of the posterior, each separated by 100 thinning samples.

For training word2vec, we set the embedding dimensionality to 200, and used a 7-word sliding window – same as the Dynamic-Eigen-Net – over the corpus (3 words flanking a middle target). We used the negative sampling optimization algorithm, with 25 negative samples, and trained the network over 40 epochs. Since word2vec initializes the weights at random, we provide results based on the average of 20 separate runs.

We used the USF norms for the free association task in order to match results from Griffiths et al. (2007). We used the same procedure that they describe to preprocess the training corpus and kept it fixed across models. That is, we used the TASA corpus and filtered out any word that either occurred less than ten times or was in a stop-list. This left us with a corpus of V = 52046 word types over D = 37650 documents; the total number of word tokens was 4402747.

With LSA and word2vec, to predict the free associates of a given cue, the cue's vector cosine with every other word was obtained and the word with the largest cosine was treated as the response. We rank-ordered the words in decreasing order, and checked the most probable human free associate's rank. A rank of one corresponds to a perfect match, whereas a rank of two means that the most common associate is below another word, and so forth.

The same procedure was applied to the Topics model by swapping cosines for probabilities. The probability of seeing a response, given a cue, is obtained by marginalizing over the topics,

$$p(response|cue) = \sum_{topic} p(response|topic)p(topic|cue)$$

In addition to the four models, we also provide results from the direct associative weights corresponding to the SPPMI normalized co-occurrence matrix (labeled as "Direct"). That is, for a given cue, the associative strengths in the corresponding row in the weight matrix were treated as activation strengths. Better performance in the Dynamic-Eigen-Net relative to the direct weights demonstrates a performance advantage gained by spreading activation.

The ranks in activation were restricted to the intersection of the words that were used as either cues or responses in the free association norms with the words in the corpus. The intersection of the word types in the USF norms and TASA was 4566.

We followed a similar procedure for applying the models to the word similarity datasets from Miller et al. (1993; MC), Bruni et al. (2013; MEN), Radinsky et al. (2011; MKTurk1), Halawi et al. (2012; MKTurk2), Luong et al. (2013; RareWord), Rubenstein et al. (1965; RG), Hill et al. (2016; SimLex), Gerz et al. (2016; SimVerb), Yang et al. (2016; YP), Finkelstein et al. (2006; WS1), and Agirre et al. (2009; WS2). For each pair of words in a given dataset, one member was used as a cue and either the resulting cosines, probabilities, or activations, depending on the model. Performance on each dataset was quantified as the Spearman's correlation between the model-derived strengths and corresponding human word similarity ratings.

Results

Figure 1 summarizes performance on the free association task across models. The left panel shows the median rank of the first associate, across all cue-response pairs, with lower valued ranks indicating that the first associate was closer to the top in strength. The Topics model, Dynamic-Eigen-Net, and the direct associations yield similar performance in terms of the median ranks (17), followed by word2vec (21.5) and LSA (29). The right panel shows the percent of times the first associate was also the most active word in each model. The Dynamic-Eigen-Net and the Topics model favour the first associate as the most active word at about the same rate, 15.73% and 15.68%, respectively. The direct associations favour the first associate 14.98% of the time, a rate lower than the Dynamic-Eigen-Net, indicating a performance advantage when spreading activation. Word2vec and LSA trail behind the other models, favouring the first associate 14.85% and 11.73% of the time, respectively.

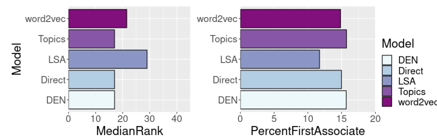

Figure 1. Performance on the free association across models shows how the Dynamic-Eigen-Net (DEN) yields low median rank for the first associate (left panel) and a high percentage of first associates as the most active word (right panel).

Figure 2 shows the Spearman's correlations between the strengths obtained for pairs of words and the human rated similarity, across models and datasets. Overall, strengths derived using word2vec show slightly higher correlations with human judgments, but the Dynamic-Eigen-Net, the Topics model, and LSA are not very far behind. The increase in the correlations for the Dynamic-Eigen-Net relative to the direct associations shows how spreading activation improves performance. The only exception is the MC dataset, which only contains 30 pairs of words. The advantage of spreading activation is most stark for the YP, SimVerb, and MKTurk1 datasets. Word2vec and LSA show superior performance to the other models for the SimLex dataset, which defines similarity based on paradigmatic relations. Paradigmatic relations hold between words that can be used interchangeably.

To explore the kinds of relations the different models capture, we obtained a set of norms where five raters categorized a set of cue-response pairs into one of six categories, including "syntagmatic", "paradigmatic", "forward", "backward", "form", and "other". The raters were asked to categorize pairs that tend to occur in the same context (e.g., WEB and SPIDER) as syntagmatic, and pairs

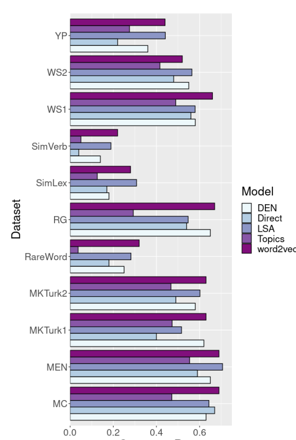

Figure 2. Performance on the word similarity datasets shows comparable performance between Dynamic-Eigen-Net (DEN) and widespread models like word2vec and Topics.

that can occur in place of one-another (TROUSERS and PANTS or EAST and WEST) as paradigmatic. When a cue (DILL) tends to precede the response in serial order (PICKLE), the raters were asked to label the pair as having a forward association and when the cue (YOLK) tends to succeed the response (EGG), it was to be labeled as a backward association. If the two words had phonetic (EYE and I) or orthographic (CHOIR and CHORE) overlap, their relation was labeled as form-based. A final category, "other", was included to make the classification exhaustive. For any given pair, each rater was asked to choose one relation. We designated the relation with the most votes across the five raters to be the dominant relation for each of the word-pairs, and probed the models with one member of the pair and tallied the median rank of the other word into bins corresponding to the dominant relation. We base the ranks on the intersection of the words in the norm-set and those in the TASA corpus, totalling 1371 words.

Figure 3 shows the median rank of the responses across word relations and models. As with the free-association ranks, lower valued ranks indicate higher activations relative to other words. The Dynamic-Eigen-Net and the Topics model show approximately the same tendency for activating syntagmatic associates, and word2vec and LSA show the least tendency. Word2vec and LSA show the strongest tendency toward activating paradigmatic relations, with the Direct associations showing the least tendency. The Dynamic-Eigen-Net shows a strong tendency toward activating words based on serial order, with a higher likelihood of activating words that succeed the cue in serial order ("forward") relative to words that precede it ("backward"). Word2vec shows the least tendency toward activating words based on serial order, indicating that such information is lost during training. Since none of the models encode form-based representations, they have no tendency towards activating other words that overlap in form.

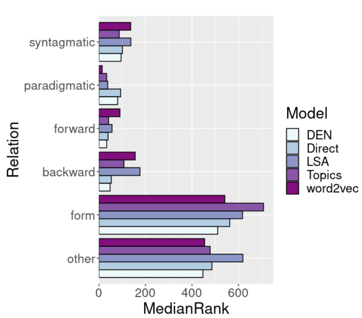

Figure 3. Median rank of the strength of associates across word relations and models.

Discussion

Our results show that assuming a latent representation may not be necessary for capturing the meaning of words. The appropriate process assumption can suffice to reveal relations between words that are not explicitly observed through the surface-level regularities in the system's record of past experience. Overall, the Dynamic-Eigen-Net outperformed word2vec and LSA on the free association task and showed a similar level of performance compared to the Topics model. Across the word similarity datasets, the Dynamic-Eigen-Net reliably outperformed the Topics model and showed a similar level of performance as word2vec on most of the datasets.

In the Dynamic-Eigen-Net, information stored in memory directly corresponds to the co-occurrence statistics of words across contexts, but the spreading of activation enables the system to induce more generic relations between words through a relaxation process. The Dynamic-Eigen-Net, being linear, facilitates the characterization of the system's dynamics based on the eigenspectrum of the weight matrix. The activation of a particular word, aligns the state vector with a subset of eigenvectors in the weight matrix and gradually shifts the state vector in their direction. Since the top eigenvectors of the weight matrix have the strongest attractive force, the state is generally pulled towards those dimensions of high variance. The high-level dynamics of spreading activation correspond to the gradual integration of the input state into the global associative structure of the entire memory system. In a similar way that LSA induces latent representations by approximating the original word-by-document matrix based on the singular vectors that capture the most variance, spreading activation drives the system's state towards the dominant eigenvectors of the weight matrix.

Deferring the construction of meaning to retrieval makes it easier to extend the system to capture the meaning of multiple words. For instance, cueing memory using the Dynamic-Eigen-Net algorithm with the word BANK, activates the words DEPOSIT, ACCOUNT, SAVINGS, CHECK, and MONEY as the top five most active words. It is possible to cue memory with more than a single word, by activating further nodes at input. Cueing memory with the two words, RIVER and BANK, activates BANKS, MISSISSIPPI, HUDSON, STREAM, and NILE as the top five most active words. The occurrence of the word RIVER in addition to BANK aligns the state of the system with a different set of eigenvectors as when cueing the network with BANK alone. The context-sensitivity of the system makes it a potential candidate for tasks like predication, without relying on extraneous processing assumptions.

Conclusion The Dynamic-Eigen-Net provides a more parsimonious alternative for capturing word meaning, relative to other accounts, and its context-sensitivity makes it promising for semantic composition.

References

Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Paşca, M., & Soroa, A. (2009). A study on similarity and relatedness using distributional and wordnet-based approaches. *In*

Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 19-27.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*, 993-1022.

Bruni, Elia, Tran, Nam-Kahn, & Baroni, Marco. (2013). Multimodal distributional semantics. *Journal of Artificial Intelligence Research, 48*, 1-47.

De Deyne, S., Navarro, D. J., Collell, G., & Perfors, A. (2021). Visual and Affective Multimodal Models of Word Meaning in *Language and Mind. Cognitive Science, 45(1)*, e1 2922.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., & Ruppin, E. (2001). *Placing search in context: the concept revisited. In Proceedings of the 10th international conference on World Wide Web*, 406-414.

Gerz, Daniela, Ivan Vulić, Felix Hill, Roi Reichart, & Korhonen, Anna. (2016). Simverb-3500: A large-scale evaluation set of verb similarity. *ArXiv preprint*.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. P*sychological review, 114(2)*, 211.

Griffiths, T., & Steyvers, M. (2003). Prediction and semantic association. *Advances in neural information processing systems*, 11-18.

Halawi, G., Dror, G., Gabrilovich, E., & Koren, Y. (2012, August). Large-scale learning of word relatedness with constraints. *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1406-1414)*.

Hill, F., Reichart, R., & Korhonen, A. (2015). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics, 41(4)*, 665-695.

Landauer, T. K., & 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(2)*, 211.

Levy, O., & Goldberg, Y. (2014, June). Dependency-based word embeddings. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 302-308)*.

Luong, M. T., Socher, R., & Manning, C. D. (2013, August). Better word representations with recursive neural networks for morphology. *In Proceedings of the seventeenth conference on computational natural language learning (pp. 104-113)*.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, George A., Charles, & Walter G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes, 6*, 1-28.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36(3)*, 402-407.

Radinsky, K., Agichtein, E., Gabrilovich, E., & Markovitch, S. (2011). A word at a time: computing word relatedness using temporal semantic analysis. *In Proceedings of the 20th international conference on World wide web*, 337-346.

Rubenstein, Herbert, & Goodenough, John, B. (1965). Contextual correlates of synonymy. C*ommunications of the ACM, 8,* 627-633.

Shabahang, K. D., Yim, H., & Dennis, S. J. (2022). Generalization at retrieval using associative networks with transient weight changes. *Computational Brain & Behavior, 5(1)*, 124-155.

Wood, J. N., & Wood, S. M. (2020). One-shot learning of view-invariant object representations in newborn chicks. *Cognition, 199*, 104192.

Yang, Dongqiang, & Powers, David, Martin. (2006). Verb similarity on the taxonomy of WordNet. *The Third International WordNet Conference: GWC 2006. Masaryk University*.

Appendix

Table A1
Small corpus

the dog played with the bone
the cat played with the leaf
the truck drove to the factory
the car drove to the garage
the flower grew on the field
the tree grew on the hill
the liberal increased taxes
the conservative decreased taxes
the eagle had white feathers
the hawk had black feathers
the bird sat along the grassy river bank
the investors stood in front of the bank
the traders deposited their money into the bank
the salesman withdrew money from the bank
the bank raised the interest rate

With V as the vocabulary size, the marginal probability of the i'th word is estimated with a as an additive smoothing parameter using,

$$p_i = \frac{C_j + \alpha V}{\sum_{l=1}^{V}[C_l + \alpha V]}$$

The marginal probability of the j'th word is estimated with a used as both the additive and multiplicative smoothing parameter using,

$$p_j = \frac{(C_j + \alpha V)^{\alpha}}{\sum_{l=1}^{V}[(C_l + \alpha V)^{\alpha}]}$$

The joint-probability of word pairs are estimated with a as an additive smoothing parameter using,

$$p_{ij} = \frac{C_{ij} + \alpha}{T + \alpha V^2}$$ where T is the total count, $$T = \sum_{i=1}^{V}\sum_{j=1}^{V} C_{ij}$$.