LayerFolding: Discovering Creative Links in Word Association Networks

Ping Xiao
Department of Computer Science and
Helsinki Institute for Information Technology HIIT
University of Helsinki
FI-00014 Helsinki, Finland
ping.xiao@cs.helsinki.fi

Hannu Toivonen
Department of Computer Science and
Helsinki Institute for Information Technology HIIT
University of Helsinki
FI-00014 Helsinki, Finland
hannu.toivonen@cs.helsinki.fi

ABSTRACT

A frequent challenge in creative tasks such as advertising is finding novel and concrete representations of abstract concepts. We cast this problem as finding, in word association networks, the relevant indirect associations of a given node. We propose a novel approach, *LayerFolding*, which selects nodes at increasing distances from the given node, according to their relatedness to it. The relatedness is calculated based on the shortest paths that are potentially coherent. In a test against a small set of visual representations of abstract concepts found in real advertisements, LayerFolding provides a 79% recall, and outperforms other two popular semantic relatedness measures.

CCS Concepts

•Information systems → Data mining; Similarity measures; •Applied computing → Arts and humanities;

Keywords

Data mining; Graph mining; Word association; Indirect association; Creativity

1. INTRODUCTION

In creative practice, a frequent challenge is using something concrete to represent an abstract concept. For instance, visual representations of 'foresight', found in print advertisements, include the pictures of eye, crystal ball, tarot card, and binoculars.

In this paper our goal is to mine word association networks for concrete concepts indirectly related to a given abstract concept. As is obvious from the above example, the abstract concept (foresight) and its concrete representations (crystal ball, binoculars, etc.) are not necessarily neighbors in the network. Actually, the opposite case might be true: a neighbor can be considered a blunt and clichéd representation, especially if the goal is to be creative. The key problem here thus is how to discover indirect yet relevant links.

Our approach of finding meaningful indirect associations takes a given node (an abstract concept) as the starting point, and iteratively explores the rest of the network at an increasing radius. We define a relatedness measure tailored to recognize remote but relevant associations. In order to filter out abstract concepts from the results, we use an existing large-vocabulary abstractness rating. We evaluate the method using a small set of visual representations found in real advertisements, and compare the results to those obtained by two generic semantic relatedness measures, Personalized PageRank [2] and Word2Vec [4].

SAC 2016, April 04 – 08, 2016, Pisa, Italy

http://dx.doi.org/10.1145/2851613.2851883

2. PROBLEM FORMULATION

Assume that a (directed or undirected) network of strong associations between words is given. *The task addressed in this paper is the following*: given a word v_a (an abstract concept) in the network, output a list of other words $v_b, v_c, \ldots$ that are relevant and novel in regard to v_a, as well as concrete.

This paper is about defining a *relevance measure* for this task and Section 3 will focus on that. In our context, "*novel*" means that a word (a concrete object) is not strongly associated to the given abstract concept, i.e. the two words are not directly connected in the given word association network. What is *concrete* is treated as a question independent of the network structure, and will be addressed with an existing abstractness vocabulary.

Consider now the word association network(s) given as input. Word associations typically are asymmetric. For instance, in free association tests, 'sheep' easily reminds people of 'dog', while 'dog' makes people first think about 'cat', 'bark', etc. The same holds for many statistical word association measures computed in text corpora; they typically estimate something like the likelihood of one word appearing in the presence of another one. These observations suggest that word association networks usually are directed, and that paths should be directed as well [8].

However, we argue that this approach may overlook some relevant associations. Consider as an example words 'foresight' and 'eye'. In graph G_{norm} of 9,680 words and 134,854 free associations, used in the experiments of this paper (see section Evaluation for details), there are eleven mixed-direction paths of three edges between them, such as $foresight \rightarrow foresee \rightarrow look \leftarrow periscope$; $foresight \leftarrow insight \rightarrow eyes \leftarrow periscope$; $foresight \rightarrow see \leftarrow sight \leftarrow periscope$. We make two observations from this. First, it seems easy for humans to interpret the connection between the end points of these paths despite their mixed directions—actually easier if directions are not given. Second, the large number of alternative mixed-direction paths actually suggests that the relation may be relevant.

Given the above observations, the methods in this paper assume an undirected, weighted graph $G = (V, E)$ where nodes in V are words and edges in E connect strongly associated words. The edge weight $w(v_a, v_b)$ of edge $\{v_a, v_b\} \in E$, for $v_a, v_b \in V$ characterizes the strength of the association.

Since many word association networks are nevertheless directed, we convert a directed network into an undirected one in the following way. Given a directed, weighted graph $G' = (V, E')$ with weight function $w'(\cdot)$, we obtain an undirected version by taking the average of the two opposite edge weights: first, $\{v_i, v_j\} \in E$ iff $(v_i, v_j) \in E'$ or $(v_j, v_i) \in E'$; further, $w(v_i, v_j) = (w'(v_i, v_j) + w'(v_j, v_i))/2$, where $w'(v_k, v_l)$ is set to zero if $(v_k, v_l) \notin E'$.

3. METHOD

We briefly outline the method here before going to the details in the subsections.

Given word v_a, the method performs a breadth-first search to other words further and further away from v_a. To formalize this, we define different layers of nodes recursively based on their distance from v_a, measured in number of edges and ignoring their weights. Let $L_a^0 = \{v_a\}$ be layer zero, i.e. the node itself. Then, for $i > 0$, layer i is defined as

$$L_a^i = \{u \in V \mid \exists w \in L_a^{i-1} : \{w, u\} \in E\} \setminus \cup_{j=0}^{i-1} L_a^j.$$

At each layer, nodes are either pruned or selected (see below). The selected nodes have two roles: they are considered to be possibly related to v_a (among which the most strongly related and concrete ones will be output); and the breadth-first search continues from the selected nodes to the next layer. This is where the name of our method, *LayerFolding*, comes from. The process stops when the outmost layer is reached. Together with the definition of relatedness below, this search method allows efficient operation over graphs. As a first step of the process we pre-compute all the layers L_a^i and record the number $|L_a^i|$ of nodes on each layer.

Measuring Relatedness.

The relatedness of v_a and v_x, denoted by rel_{ax}, is calculated based on the set P_{ax} of shortest paths found between them. Let v_x be on the ith layer. Given the breadth-first search method, each shortest path from v_a to v_x passes through a selected node at layer $i-1$. Now, instead of explicitly considering all shorted paths, we take advantage of the relatedness values already computed for those nodes at layer $i-1$. Let v_y be such a node at layer $i-1$ via which the shortest path traverses. We define the *coherence* $Co(v_a, v_y, v_x)$ of the connection between v_a and v_x via v_y as the minimum between the relatedness of v_y to v_a, and the weight of the edge connecting v_y and v_x: $Co(v_a, v_y, v_x) = \min(rel_{ay}, w(v_y, v_x))$.

The motivation for this choice is that the coherence of a path, i.e. to which extent the nodes at the two ends are related, partially depends on where it is most likely to go awry – the weakest link. Since our intention is to find novel, i.e. distant associations of v_a, we do not separately penalize paths for their length. Note that using the breadth-first search, coherence can be computed at a constant extra cost as part of the search.

We assume that nodes which have more coherent connections are more likely to be related to v_a. Relatedness rel_{ax} is therefore based on the sum of coherence $Co(v_a, v_y, v_x)$ of connections over all the selected nodes v_y at layer $i-1$ that have an edge to v_x. However, the number of connections can increase with the length, e.g., nodes further away from v_a have more connections than the neighbors (which only have one). Consequently, we apply a scaling factor $1/\alpha_{ax}$ (defined below) to the sum of coherence, in order to counterbalance this effect: $rel_{ax} = \dfrac{\sum_{\{v_y, v_x\} \in E} Co(v_a, v_y, v_x)}{\alpha_{ax}}$.

Scaling Factors for Longer Shortest Paths.

Let d_a and d_x be the degrees of v_a and v_x respectively, and N the number of nodes in the network. We approximate the probability that v_a and v_x are neighbors by $p_1 = \dfrac{\max(d_a, d_x)}{N-1}$.

If v_x is not a neighbor of v_a, the probability that v_x connects to at least one of v_a's neighbors, i.e. v_x is a layer-2-node of v_a, is $p_2 = 1 - \dfrac{\binom{N-d_a-2}{d_x}}{\binom{N-2}{d_x}}$, where the divident is the number of ways of choosing the d_x neighbors of v_x from all nodes (N) except any of the d_a neighbors of v_a, and except nodes v_a and v_x. The denominator is the number of ways of choosing the d_x nodes among any of the possible nodes (which, again, do not include v_a and v_x).

Since N is reasonably large — approximately 10,000 nodes in G_{norm} — and d_a and d_x are relatively small — the average degree in G_{norm} is about 24 — good approximations of p_2 can be obtained as $p_2 \approx 1 - (\frac{N-d_a-2}{N-2})^{d_x}$ (cf. sampling without vs. with replacement).

We use these probabilities to estimate the scale of change in the number of shortest paths, due to random effects, as the shortest distance between v_a and v_x increases. Assume that node v_x is on layer i for some $i > 2$. Let $sel(j)$ denote the number of nodes selected from layer j. We want to estimate the probability p_i that v_x is connected to at least one of the $sel(i-1)$ selected nodes on the previous layer. Given that it is on layer i we know that it is by definition not connected to any layers $j \in [0, i-2]$. The probability is then obtained,

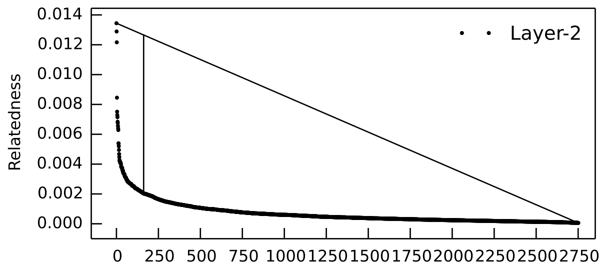

Figure 1: Distribution of relatedness between the concept 'strong' and its layer-2-nodes.

similar to p_2, as

$$p_i = 1 - \frac{\binom{N-\sum_{j=0}^{i-2}|L_a^j|-sel(i-1)-1}{d_x}}{\binom{N-\sum_{j=0}^{i-2}|L_a^j|-1}{d_x}} \tag{1}$$

$$\approx 1 - \Big(\frac{N-\sum_{j=0}^{i-2}|L_a^j| - sel(i-1) - 1}{N-\sum_{j=0}^{i-2}|L_a^j| - 1}\Big)^{d_x}. \tag{2}$$

The scaling factor is then in general obtained as $\alpha_{ax} = \dfrac{p_i}{p_1}$.

Selection of Related Associations.

An example of the relatedness between v_a = 'strong' and the nodes in one of its layers, as produced by LayerFolding, is shown in Figure 1, where the relatedness values are sorted from high to low. The distribution clearly has a long tail and an elbow. We take advantage of the elbow to select the nodes more related to v_a within each layer.

To decide the turning point of an elbow, we draw a straight line between the beginning and end points of a curve, and then select the point on the curve which is the furthest from the straight line (as illustrated by the two straight lines in Figure 1, where the intersection does not look like a right angle because of the different scalings of the two axes).

At the end of the iteration, the selected associations in every layer are merged into a single list and sorted in the descending order of relatedness. Concrete concepts are extracted from this list, as output, using an existing list of abstractness ratings for 114,501 terms [6].

4. EVALUATION

We evaluate the LayerFolding approach by testing how well existing visual representations of abstract concepts can be rediscovered by this method. We also compare the results to those obtained by two common relatedness measures. Before presenting the details of the evaluation, we first introduce the word association network used in this work.

4.1 Word Association Network

Word association norms are natural resources of relations between concepts. They are acquired from human subjects experimentally by asking them to provide the first word that comes to their mind after a cue word is presented to them. To construct a network based on this data, the cue and response words are the nodes. There is a directed edge from each cue word to each of its response words. The edge weight is set to the conditional probability of the response word given the cue word. We combined the two largest word association norms we could obtain, the Edinburgh Associative Thesaurus (EAT) [3] and the University of South Florida Free Association Norms [5]. Idiosyncratic responses — those produced by one person only — were discarded, and so were multiple arcs and self loops. Also, only normed words are included, i.e. we excluded response words that were not used as cues, too. There are 18,276 overlapping edges between the two sources. Since they are obtained with the same method, we merged each pair of parallel edges and used their weighted mean as the edge weight. The resulting network, denoted as G_{norm}, has 9,680 nodes and 134,854 directed edges. Due to the experimental setting, the response words are generally considered as strong associates of the cue words. Therefore, there is no need to select in the layer 1 of a G_{norm} node.

4.2 Recall of Visual Representations Used in Advertisements

We obtain our test cases from a data set of 37 distinct visual (concrete) representations found in real advertisements, representing six abstract concepts [7]. We ignored those abstract concepts and visual representations that are not included in G_{norm}, obtaining four abstract concepts and a total of 19 visual representations of them (Table 1, first and second columns respectively). Out of the 19 representations, 11 are immediate neighbors of the respective abstract concept in G_{norm}, and eight are non-neighbors (words in bold in the second column of Table 1).

In the test, the four abstract concepts were given as input to LayerFolding. LayerFolding then discovered for each input concept possible concrete representations, and ranked them using rel_{ax} defined in the previous section. The rankings of the known 19 representations of the four concepts are listed in the third column of Table 1. The four "N/As" indicate that the corresponding visual representations were not found by LayerFolding. In total, LayerFolding covers 79% of the 19 visual representations. Interestingly, LayerFolding was able to find four of the eight non-neighbors (in bold), and in the case of "foresight" even gave them high ranks.

4.3 Comparison to Other Measures

We compare the rankings provided by LayerFolding with the ones computed by Personalized PageRank (PPR) [2] and Word2Vec [4]. PPR is a popular measure of relatedness in networks, and it has been previously applied on a graph derived from Wikipedia in order to discover interesting and surprising links between entities [1]. Word2Vec is a state-of-the-art measure of semantic similarity, and known for facilitating creative analogy making [4].

We applied both PPR (in a similar way as in [2]) and Word2Vec to obtain the relatedness/similarity between each of the four abstract concepts and each of the concrete concepts in G_{norm}, i.e. applying the same abstractness vocabulary we use in LayerFolding. Then, the concrete concepts were ranked based on the obtained relatedness/similarity

Table 1: The visual representations of four abstract concepts ranked by LayerFolding, Personalized PageRank and Word2Vec. Visual representations in bold are non-neighbors of the respective abstract concepts.

Abstract concept	Visual represent.	Layer-Folding	Personal. PageRank	Word 2Vec
strong	**tank**	N/A	1,013	1,086
	ox	30	88	372
	rope	N/A	990	653
	muscle	2	2	49
soft	cloth	45	91	208
	sofa	40	72	486
	towel	165	1,278	698
	pillow	2	3	42
	teddy bear	16	33	N/A
	baby	31	124	1,913
	butter	109	358	14
	cotton	5	12	206
intelligence	**owl**	N/A	232	2,075
	book	16	20	2,129
	brain	1	1	95
	chess	N/A	1,267	341
	genius	2	2	132
foresight	**eye**	3	3	1,493
	periscope	8	3,146	1,038
average		*32*	*135*	*461*

values (the fourth and fifth columns of Table 1, respectively).

The magnitudes of rankings given by the three methods clearly show that LayerFolding is the most accurate in the sense of ranking known visual representations closer to the top. For 14 out of the 15 representations that it finds, it gives the best ranking among the three methods.

For easier comparison between the methods, we also computed average ranks for each method over the top 15 ranks it gave among the 19 visual representations (bottom row of Table 1). We ignored four worst results for each method, to make handling of "N/As" easy (they are simply ignored) but still keeping the comparison fair. These average rankings obtained by LayerFolding, PPR, and Word2Vec are 32, 135, and 461, respectively.

5. CONCLUSIONS

We have proposed *LayerFolding*, a method of finding creative associations of a given concept in a network of word associations. The method is motivated by the common creative task of representing abstract concepts with concrete things, where relevance and novelty need to be balanced.

LayerFolding is essentially a graph-based relatedness measure. It addresses the issue of relevance with a recursively defined measure of coherence of the paths between two nodes. This measure then allows the method to discover associations that are novel in the sense that they connect words that are not connected with an edge.

In an empirical test using a small set of visual representations of abstract concepts found in real advertisements, LayerFolding provided 79% recall and clearly outperformed two popular semantic relatedness measures, Personalized PageRank and Word2Vec. This comparison shows that the problem of finding creative links is different from measuring semantic relatedness/similarity, and methods for the latter are not effective in this task. Instead, novel methods such as LayerFolding are needed.

Evaluation of creative tasks like this one is difficult due to the lack of labeled datasets against which to assess their performance. The tests of this paper evaluate the recall of the methods with respect to a small set of known concept–representation pairs. In the future, we plan to evaluate LayerFolding with user studies, in order to get a fuller picture of its performance, including precision. We also consider applying LayerFolding to other word association networks to assess the generality of the method.

6. ACKNOWLEDGMENTS

This work has been supported by the European Commission (FET grant 611733, ConCreTe) and by the Academy of Finland (decision 276897, CLiC).

7. REFERENCES

[1] C.-Y. Chang, S. Clark, and B. Harrington. Getting creative with semantic similarity. In *IEEE 7th International Conference on Semantic Computing*, ICSC '13, pages 330–333, 2013.

[2] T. H. Haveliwala. Topic-sensitive PageRank. In *Proceedings of the 11th International Conference on World Wide Web*, WWW '02, pages 517–526.

[3] G. R. Kiss, C. Armstrong, R. Milroy, and J. Piper. An associative thesaurus of English and its computer analysis. In A. J. Aitkin, R. W. Bailey, and N. Hamilton-Smith, editors, *The Computer and Literary Studies*, pages 153–165. 1973.

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. 2013.

[5] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber. The University of South Florida Word Association, Rhyme and Word Fragment Norms. http://web.usf.edu/FreeAssociation/Intro.html. Accessed: 2015-08-20.

[6] P. D. Turney, Y. Neuman, D. Assaf, and Y. Cohen. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 680–690, 2011.

[7] P. Xiao. *Generating Ideas for Pictorial Advertisements: Starting with Pictorial Metaphors*. PhD thesis, Universitat Pompeu Fabra, 2014.

[8] K. Zhang, K. Zhu, and S.-w. Hwang. An association network for computing semantic relatedness. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, AAAI '15, 2015.