cite2vec: Citation-Driven Document Exploration via Word Embeddings

Matthew Berger, Katherine McDonough, and Lee M. Seversky

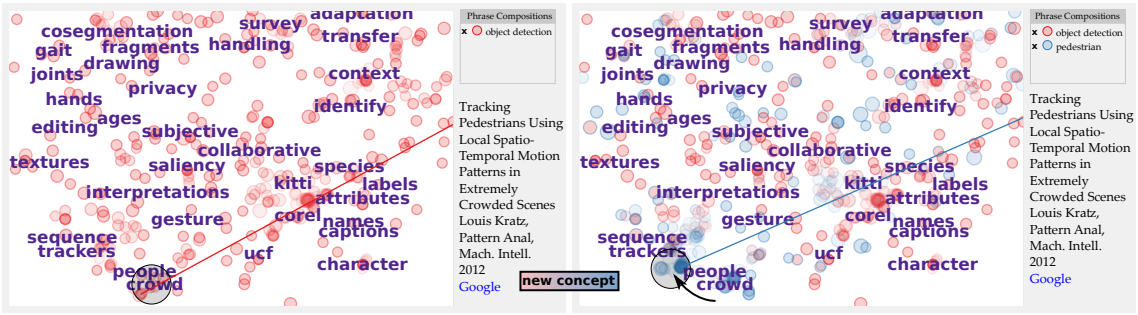


Fig. 1. cite2vec enables exploration of document collections via *citation contexts*: how people tend to use documents. Our interface enables document usage exploration via user-prescribed concepts: here we show a 2D projection of words, where each word is composed with the specified phrase "object detection" (left). Each document – shown as a disk – is positioned near words whose phrase compositions best describe its usage, here taken as "crowd object detection". The user can add phrases (right) to compare usage, resulting in new document projections, with the highlighted document taking on the more precise usage "pedestrian trackers".

Abstract— Effectively exploring and browsing document collections is a fundamental problem in visualization. Traditionally, document visualization is based on a data model that represents each document as the set of its comprised words, effectively characterizing what the document is. In this paper we take an alternative perspective: motivated by the manner in which users search documents in the research process, we aim to visualize documents via their *usage*, or how documents tend to be used. We present a new visualization scheme – cite2vec – that allows the user to dynamically explore and browse documents via how other documents use them, information that we capture through *citation contexts* in a document collection. Starting from a usage-oriented word-document 2D projection, the user can dynamically steer document projections by prescribing semantic concepts, both in the form of phrase/document compositions and document:phrase analogies, enabling the exploration and comparison of documents by their use. The user interactions are enabled by a joint representation of words and documents in a common high-dimensional embedding space where user-specified concepts correspond to linear operations of word and document vectors. Our case studies, centered around a large document corpus of computer vision research papers, highlight the potential for usage-based document visualization.

Index Terms—document visualization, word embeddings

1 Introduction

Document collections have been growing at a rapid pace in recent years. The massive volume of documents in such areas as scientific literature, news articles, and social media has necessitated visualization techniques that enable user tasks such as understanding latent themes in a corpus, seamlessly exploring document relationships, and discovering documents of interest. In these forms of analysis a document is traditionally modeled by its underlying content, and in particular, the set of words comprising the document. This can be seen in visualization schemes that explore documents through 2D projections [11, 38, 42], or those that convey themes in a corpus via topic models [15, 16, 27, 41].

These tools are built around a document model that characterizes *what* a document is. A document can be represented in many different ways, however, and a content-based document model may be unsuitable for certain tasks. For users who need to explore and discover documents as part of the research process, such a model may fail to capture the aspects of a document that are useful for this process – in particular, the common usage of a document, and *how* a document is used. As demonstrated in previous studies [6, 19], users tend to seek out documents for rather diverse uses – for instance in [6], journalists often searched documents to provide evidence supporting their arguments and overall story. In scientific literature, researchers often seek papers used in specific ways, for instance as a dataset, benchmark, or evaluation criteria. Document visualization based on usage thus has the potential to greatly benefit users in the research process.

In this paper we introduce cite2vec – a method for exploring and discovering document collections via document usage. To capture document usage, we observe that the *context* in which one document cites another tends to reflect how a document is used, namely, within a document, people tend to cite other documents for very precise reasons. cite2vec visualizes documents and words in a way that adheres to such citation contexts, so that the user can explore and discover documents in a usage-oriented manner. Our approach projects words that are representative of documents into a 2D space, and subsequently projects documents into this same space such that a document's proximity to a word indicates its manner of usage, while also preserving similarity of document-to-document usage. More importantly, we allow the user to interactively modify the document projections to their interests by treating the 2D space as a way to specify *concepts*. Namely, the user can specify arbitrary phrases that change the meaning of the projected words: each word is *composed* with any of these phrases, or pairwise phrase combinations, with documents positioned and encoded with the word-composition that best characterizes its usage, as illustrated

- *Matthew Berger and Lee M. Seversky are with the Air Force Research Laboratory.*
- *Katherine McDonough is with Northeastern University.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

in Figure 1 for an illustration. We also allow for the user to specify documents themselves for composition, mixing words and documents to best explore the space. Additionally, we allow for the 2D space to take on the concept of *analogies*, where upon specifying a "document-is-to-word" analogy, the 2D concept space represents potential analogy completions. All forms of interaction dynamically filter documents and modify document positions so that they best reflect the user-specified concepts, allowing the user to compare documents by different forms of usage, identify trends in concepts, and steer document exploration through usage-based interaction.

To enable these tasks it is essential to have a model that integrates words, and more generally arbitrary phrases, with documents, such that document usage is faithfully represented, word representations are semantically meaningful, and the above user interactions are naturally supported. We achieve this by embedding both words and documents into a common high-dimensional space by extending the powerful word2vec model [30], originally designed to learn embedding spaces which preserve word semantics. We show how to treat documents as words via their *citations*, allowing us to jointly embed words and documents in the same space. This allows us to capture document usage by modeling a document as being predictive of the words surrounding all of its citations. Notably, the embeddings in word2vec have a demonstrated *linear structure* – the addition of two words' vectors results in a new vector whose closest word is semantically similar to the composition of the words. This property enables the operations of phrase/document composition and analogy specification, and drives the dynamic exploration of documents.

To summarize, our two main contributions are a novel representation of documents and words and how to use this for visualization:

- Data representation: We learn an embedding space of words and documents that exhibits linear structure, permitting the semantically meaningful composition of words and documents.
- Visual interaction: We dynamically explore documents via user-tailored concepts, enabled by the data representation.

We demonstrate the effectiveness of cite2vec by considering the exploration of a large corpus of computer vision research papers. We detail a case study of an experienced machine learning researcher who is investigating an unfamiliar subdiscipline of computer vision, illustrating how our system allows the discovery of different types of research via their usage. Secondly, we attended a computer vision conference to demo our system to experts in the field, obtaining their feedback on our system. Our case studies highlight the potential of cite2vec.

2 Related Work

As our work is a citation context-driven manner of exploring documents via 2D projections, we span two basic forms of visualization research: document visualization, with an emphasis on 2D dimensionality reduction techniques, as well as citation network visualization. We discuss these related works below.

2.1 Visualizing Document Collections

A traditional approach to visualizing documents is through projecting each document into a 2D space, where documents that are similar in some sense are positioned close to each other in 2D. Document similarity is typically defined via distances of a document-term vector space representation: each document is represented as a point in a high dimensional space where each dimension is a term, and each term's contribution is proportional to its frequency in the document – often weighted via the term frequency-inverse document frequency (tf-idf) scheme. Dimensionality reduction schemes built on this representation have been used for 2D document projection such as multidimensional scaling [42] and self-organizing maps [38]. Recently, t-SNE [40] has proven a very effective means of visualizing structure in high-dimensional data via a heavy-tailed probabilistic model, used in document visualization in [9, 27], with recent extensions focused on exploiting sparsity in document-term representations [23].

A single 2D projection of documents, however, may fail to capture the multiple forms of similarities between documents. Hence, interaction is necessary to assist the user in effectively exploring a document collection. A common form of interaction is to allow the user to manually position a small subset of documents into a 2D space, followed by projecting the remainder of the documents. This can be seen in the Least Squares Projection approach [33], which preserves document similarity via Laplacian regularization, constrained by a small number of documents used as control points in the 2D space. Subsequent approaches have extended this methodology, using hierarchical sampling [32], locally-affine projections [24], and radial basis function regression [2]. Such interactions, however, are inappropriate for document discovery, as the user must know apriori how documents relate in order to properly specify their 2D positions. The primary issue is the representation: as our approach represents documents and words in the same space, we may reason about documents via arbitrary word phrases – this drives the dynamic projection of documents, providing for more intuitive user interaction in discovering documents.

Topic models – in particular Latent Dirichlet Allocation (LDA) [5] seek a more semantic and explainable representation of documents. Here, a document collection – a document-term matrix – is represented by a generative process that models each document as a distribution over latent topics, and each topic is modeled as a distribution over words. The topics capture informative yet discriminative themes that correlate well with human interpretation [7]. Many visualization systems have been built around topic models: TIARA [41] depicts time-varying topic evolution via the ThemeRiver [20] metaphor, ParallelTopics [15] treats topics as ordered dimensions in a parallel coordinates visualization, HierarchicalTopics [16] performs agglomerative clustering on topics in order to scale to large numbers of topics, and [31] visually compares different document collections via topic (dis)similarities.

Interaction in topic models is mostly limited to exploring the learned models, with different views of documents largely driven by the assigned topic distributions. Hence one major difference with our approach is that we enable the user to explore documents through arbitrary word phrases, rather than with fixed topic distributions. This gives the user flexibility in exploring the space – owing to our model's capability of preserving word semantics – instead of being tied to the given set of topics' words. It is possible to adjust the parameters in LDA to allow for more expressive models, yet the number of topics and representative words per topic have been shown to be difficult to set for visualization purposes [16, 41]. Several approaches [9, 27] permit user interaction in topic models to modify the underlying model optimization, but these approaches change document relatedness indirectly via model modifications, rather than using the model directly.

2.2 Citation Networks

All of the above approaches share the same form of document-term representation, summarizing what a document is. Our approach instead models a document from its usage via citation contexts. The structure of document citations has been an active visualization area, used to better understand trends in document collections, to assign a notion of importance to certain documents, as well as understand topic lineage. There have been many efforts to use citation networks to compute statistics on document collections, such as the EigenFactor metrics [4], which uses eigenvector centrality to rank journals based on citations and journal importance.

These types of statistics enable the visualization of documents through their citations. Citespace II [8] allows the user to visualize documents via citation network graphs under different time windows, as well as citation histories of articles. The analytics tool of [17] allows one to explore a citation network via graph visualization of the network, in addition to various network statistics, as well as the citation contexts themselves. Metro maps [37] attempts to detect and visualize dependency structures in citation networks, in order to provide the user a lineage of documents for better understanding. Co-citation analysis is performed in [43] to depict hierarchical relationships between papers. Citation networks have also been used as part of visualizing literature surveys [3], in order to visualize various statistics for a research area.

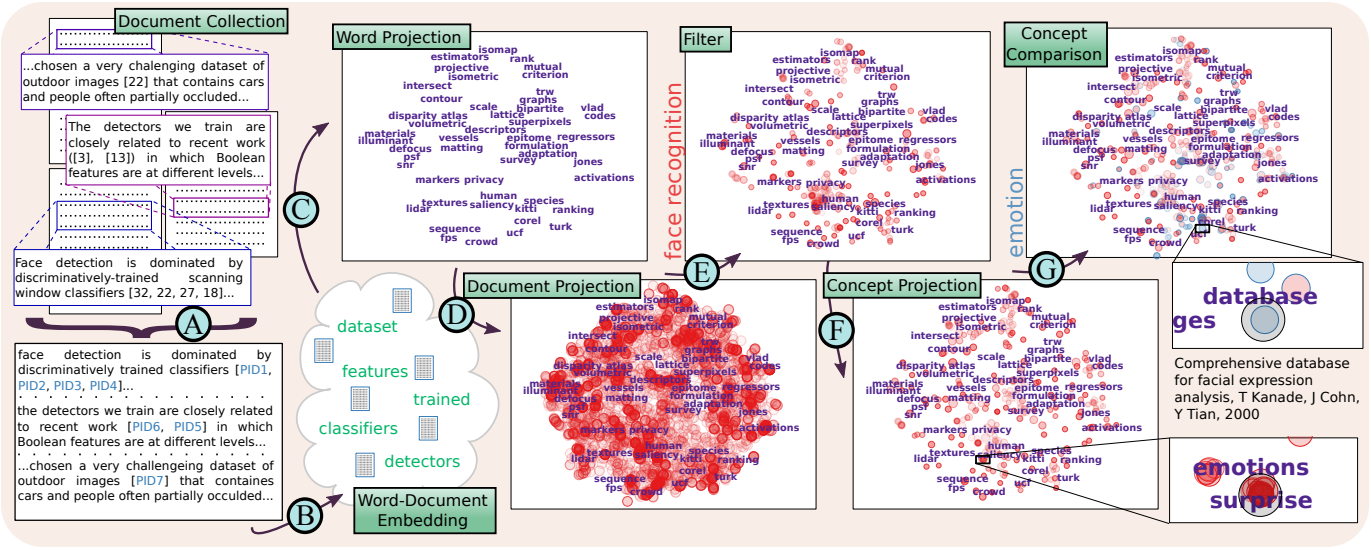


Fig. 2. Overview of cite2vec: from a given document collection we resolve all document citations to unique identifiers, and jointly learn a semantic embedding of words and documents. We then perform a 2D projection of the words and documents from the embedding, and allow the steering of document projections via user-defined concepts. Here, the user specifies "face recognition", resulting in each word being composed with this provided phrase. Upon specifying "emotion" we observe a document change its projection due to a better word-phrase composition.

These approaches, however, tend to not use the actual document content as part of their visualization. A notable exception is the CiteRivers system [21], which uses citation networks in conjunction with the actual document content to visualize various aspects of document collections: prominent journals, author evolution over time, and trendiness in documents. However, since CiteRivers models documents through their underlying terms, it fails to take into account the rationale of document citations. As a result, the representative use of documents is not captured.

3 CITE2VEC OVERVIEW

cite2vec is largely motivated by an important rationale for users to research document collections: find documents to use for a specific purpose. We aim for document visualization that encodes such usage, so that users can discover documents by how other people have used them. With this in mind, we aim to support three basic user tasks:

T1: Present usage overview. The user should be able to comprehend themes in document usage to inspire more detailed explorations.

T2: Steer document exploration via usage. The language and semantics by which people use documents should be the basis in how a user interacts with documents.

T3: Compare document usage. The user should be able to discover similarities/differences in how documents are used.

Figure 2 provides an overview of cite2vec, illustrating how we achieve these tasks. The first step (A) of our method is to gather a collection of documents, and resolve citations: a given document's citations across all documents is replaced with a unique id. We aggregate all documents into a single sequence of text and learn an embedding over both words and documents (B), treating each document as a unique word in the learning procedure. The embedding preserves word semantics, word-document usage relationships, and inherits a linear structure, all of which inform our visualization methodology.

In our visual interface, we first project (C) document-relevant words from the high-dimensional space into 2D in a multi-scale manner so that as the user zooms in, they can observe more detailed concepts (**T1**). We next project (D) the documents, rendered as disks, into the space, satisfying two objectives: documents remain close to relevant words, and document-document similarity is preserved. We then utilize the linear structure of the word embeddings to enable the user to modify the document projections to their interests (**T2**). Upon specifying a word phrase, each word in the 2D space is composed with the phrase, and documents are positioned near words whose compositions best reflect the document usage: first irrelevant documents are filtered (E), and the relevant ones then move to their best composition (F), while still respecting document structure. Note the cluster of documents near human (E) spread out upon specifying "face recognition" (F), as these documents are better described by more precise compositions with this phrase. We allow the user to specify multiple phrases (G), such that each word can take on the meaning of any one phrase or pairwise combination of phrases, enabling the user to compare documents via different types of usage (**T3**). Note that the particular document shown changes its position upon specifying "emotion", as "emotion database" best reflects how the paper is used.

3.1 Dataset: Computer Vision Research Papers

In order to motivate and highlight our approach in subsequent sections, we first describe the dataset we have used for our experiments. Our document collection consists of the space of computer vision research papers. Computer vision papers are a very rich document corpus to explore usage. Research papers can be characterized in terms of subdisciplines such as categorization, segmentation, and reconstruction, different forms of visual data such as images, video, human pose, and trajectories, techniques spanning heterogeneous research domains such as optimization and machine learning, and other forms of use ranging from datasets, evaluation, benchmarks, and code. The rapid pace of the field thus necessitates researchers to discover documents for specific reasons.

To form our dataset we have collected research papers from the following computer vision conferences and associated years: IEEE Conference on Computer Vision and Pattern Recognition (2003-2015), IEEE International Conference on Computer Vision (2003-2015), European Conference on Computer Vision (2000-2014), British Machine Vision Conference (2003-2015), and IEEE Winter Conference on Applications of Computer Vision (2007-2016). In total, we collected 12,223 papers as pdfs, and for each we extract the raw text via a standard pdf to text parser. We then employ ParsCit [10] on each text document to extract its title, authors, year, the main text body, paper references, and citations. As paper references are often slightly different across documents, due to such factors as typos, differing authors, title abbreviations, or errors in ParsCit, we perform identity resolution by resolving papers if the string edit distance of their titles is within a threshold, or if the author last names and year of publication match. Though we may erroneously identify different papers as the same, we have not found this noise to have an impact on our approach. Last, we process

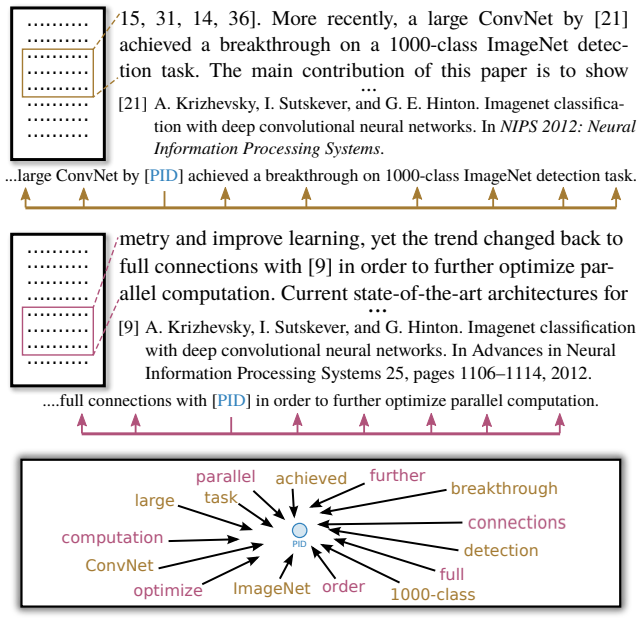


Fig. 3. Illustration of the Skip-gram model for citations. The document denoted by PID is referenced by two different documents (top and middle), highlighted by its citation contexts in both documents. These contextual words become associated with the document in the embedding space (bottom).

each document's main body and replace each citation with a unique id, so that different documents which cite the same document are now properly referencing the same id.

The output of this process is the concatenation of all documents' main bodies, represented as one large stream of text. Note that we discard the original document information, i.e. each sentence in the dataset is no longer identified with its corresponding document. Instead, the documents we aim to visualize are those which are cited in the corpus – in our dataset, this amounts to a total of 71,314 documents. Our objective is to learn a model over these documents via *how they are cited* through surrounding citation text, in addition to learning a model over all of the words in the corpus, so that we can build flexible visualization schemes for exploring documents and their common usage. We do so by learning an embedding over words and documents, which we detail in the next section.

4 Jointly Learning and Projecting Word-Document Embeddings

To model citation contexts, we wish to jointly learn an embedding over words and documents so that words and/or citations which co-occur in sequential text neighborhoods remain close in the embedding. This provides for a means of similarity between words, documents, and word-document pairings. To this end, we employ the word2vec model [30] in order to find such an embedding, a neural language model which supports training from large amounts of text data. We first provide an overview of word2vec's Skip-gram model, and then illustrate how we extend it to model documents via their citations.

4.1 Skip-gram Model

The Skip-gram model takes in a large sequence of words and assigns a point in a high-dimensional space to each word, such that other words in local neighborhood contexts remain close to this word in the embedding space. The intuition is that each occurrence of a given word should be predictive of the surrounding words in the text. More specifically, suppose we represent our text data as a sequence of words $W = (w_1, w_2, ..., w_n)$ with each word belonging to a pre-defined vocabulary V_w of size n_w. We associate word $w \in V_w$ in the vocabulary with a *word vector* in a d-dimensional space $\mathbf{x}_w \in \mathbb{R}^d$ – this represents our word embedding. We also associate a *context vector* $\hat{\mathbf{x}}_w \in \mathbb{R}^d$ for word

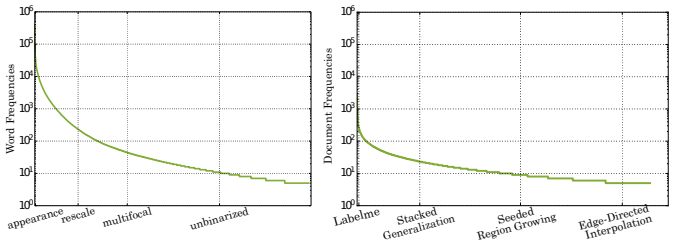


Fig. 4. We show unigram distributions for words (left) and documents (right), sorted by frequencies, showing representative words and documents from their vocabularies of size 44,752 and 18,832, respectively. The discrepancies between their distributions necessitate a different negative sampling scheme in the skip-gram model.

w. We would like a given word vector to be predictive of its surrounding words' context vectors, but not necessarily the word vectors – i.e. if the context and word vectors were the same, then a word would be most predictive of itself, however it is uncommon for a given word to appear in its own surrounding words.

The Skip-gram model aims to find an embedding such that words which co-occur are close in the embedding, while also ensuring that a word is far from other words randomly drawn from some distribution P, a process known as *negative sampling*. The distribution P is typically taken as the unigram distribution over V_w (normalized word counts), raised to a certain power, $\frac{3}{4}$ in [30]. The objective of the Skip-gram model is to maximize the following objective with respect to all word and context vectors:

$$E = \sum_{i=1}^{n_w} \sum_{-c_w \le c \le c_w, c \neq 0} \left(\log(\sigma(\mathbf{x}_{w_i}^T \hat{\mathbf{x}}_{w_{i+c}})) + N(\mathbf{x}_{w_i}) \right), \quad \text{(1a)}$$

$$N(\mathbf{x}) = \sum_{j=1}^{k} \mathbb{E}_{w \sim P} \left(1 - \log(\sigma(\mathbf{x}^T \hat{\mathbf{x}}_w)) \right), \quad \text{(1b)}$$

where c_w is the context window size, k is how many negative samples to draw per word, and σ is the sigmoid function $\sigma(x) = \frac{1}{1+\exp(-x)}$. The window c_w determines how many surrounding context vectors should be similar to the given word vector, while k determines how many random context vectors should be dissimilar. Intuitively, the first term encourages word-contexts that co-occur to have high likelihood, and are consequently close in the embedding, while the second term N encourages words to be sufficiently distinct from randomly drawn contexts, and consequently far apart in the embedding.

4.2 Citations as Words

The Skip-gram model, as defined above, scans over the text corpus to build the vocabulary V_w. We augment this model with a new vocabulary V_d of n_d documents which are cited by the document collection, in order to learn an embedding over words and documents in the same embedding space. We associate document $d \in V_d$ with both a vector in the embedding space $\mathbf{z}_d \in \mathbb{R}^d$, and a context vector $\hat{\mathbf{z}}_d \in \mathbb{R}^d$. We may then run the Skip-gram model, as defined in Equation 1a, on the corpus with respect to both V_w and V_d, simply treating citations as unique words. Figure 3 illustrates this process, showing how citations from different documents can result in diverse word associations. The words surrounding all citations of a given document serve to uniquely characterize the document by bringing these words closer to the document in the embedding space, hence capturing the common usage of documents. Note that this goes both ways: words become more distinct via their associations with citations.

An issue with the original word2vec model for our setting is the disparity between the separate word and document unigram distributions. Figure 4 shows the histograms for words and documents in our computer vision document dataset, highlighting the substantial differences. The primary issue lies in negative sampling: documents

are rarely used as negatives, hence words are not discriminative with respect to documents. To rectify this, we sample uniformly over the word unigram and document unigram distributions, denoted P_w and P_d, respectively. We modify the negative sampling N from Equation 1b as follows:

$$N(\mathbf{x}) = \sum_{j=1}^{k} \left(\mathbb{E}_{w \sim P_w} \left(1 - \log(\sigma(\mathbf{x}^T \hat{\mathbf{x}}_w)) \right) + \mathbb{E}_{d \sim P_d} \left(1 - \log(\sigma(\mathbf{x}^T \hat{\mathbf{z}}_d)) \right) \right) \tag{2}$$

4.2.1 Implementation Details

In practice, we solve Equation 1a via stochastic gradient descent, extending the word2vec implementation[1]. We set the context window c_w to a fixed amount for words. The setting of c_w is particularly important: smaller window sizes capture syntactic regularities, while larger window sizes forego this for broader, more semantic meanings [28]. For our dataset, the linguistics of computer vision research papers can be quite dense, hence to adequately capture the underlying semantics we require a somewhat large window size, which we set to 30 in our experiments, though setting c_w to 30 ± 10 produced similar results.

For citations we dynamically adjust c_w so that it expands to the preceding, current, and subsequent sentences surrounding the citation, similarly to [22]. This is based on the observation that a citation tends to have high relevance to words in a large context. We also take care to handle a block of citations, such that if one belongs to the neighborhood context of a given word, it is considered a single word separated into its constituent documents.

Following [30] we discard infrequent words and documents from the corpus, setting the minimum threshold occurrence to 5 for both, since it is challenging to find adequate representations with fewer word/document frequencies. This results in vocabulary sizes of $44,752$ and $18,832$ for words and documents in our dataset, respectively. We set the dimensionality of the embedding d to 150, though we found embedding qualities to be comparable for $d = 150 \pm 25$. The negative sampling size k is set to 6 following [30] for sampling from both the document and word unigram distributions.

4.3 Embedding Properties

The learned embeddings of the words and documents exhibit several useful characteristics. First, the learned word vectors carry the semantics of the underlying words. We highlight this in Figure 5, where we perform k-means on the word vectors for $k = 300$, and for each cluster we show the 5 words with highest occurrence frequency over the corpus. Similar to document-oriented models like LDA, some of the clusters correspond to salient topics, such as tree structures, segmentation characteristics, and facial features. However, the clusters encompass a wider variety of semantics: authors, implementation use, human activities, and lower-level semantics such as size and basic geometry concepts. These fine-grained semantics are essential in enabling the user to meaningfully explore the embedding space, and extend to the learned document vectors as well.

Furthermore, the embedding space inherits a linear structure [30]: words in a given context neighborhood are in linear relationship with each other with respect to the sigmoid function, in that the sum of two word vectors is proportional to their log-product. Since the training objective is to predict words in a local neighborhood, this implies that the addition of two word vectors should remain close to any other word in the neighborhood. This reasoning extends to document vectors, enabling us to relate words and documents via simple arithmetic. This allows the user to express arbitrary phrases via simple addition of the set of words, in addition to document-word *analogies*, i.e. for words w_a, w_b and documents d_a, d_b, we may express the analogy $d_a : w_a$ as $d_b : w_b$ as satisfying $\mathbf{z}_a - \mathbf{x}_a \approx \mathbf{z}_b - \mathbf{x}_b$. The embedding's semantic and linear properties support useful techniques for performing 2D projections of documents and words, so that the user can interact with the embedding in an expressive manner.

[1] https://code.google.com/archive/p/word2vec

Fig. 5. We show clusters of words associated with the learned cite2vec embedding. Note the diversity in the clusters: some refer to standard computer vision topics, while others are more general semantic concepts like animals, climate, and funding acknowledgments.

4.4 Word Projections

Given the embedding, we next wish to find a 2D projection over a careful sampling of words. We sample words in such a way that:

- The words cover general themes found within documents.
- Words are semantically distinct from one another.
- The sampling of words is done in a multi-scale manner.

Notice that a straightforward sampling scheme such as k-means over words can produce a diverse sampling, but the words may not be representative of document usage. Instead, we utilize the joint document-word embeddings to achieve these objectives by sampling words through their proximity to documents.

More specifically, we first perform farthest point sampling of document vectors. Namely, assuming that we have sampled i document vectors $\mathbf{Z}_i = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_i)$, and given all document vectors $\mathbf{Z}$, the $i+1$ document $\mathbf{z}_{i+1} \in \mathbf{Z}/\mathbf{Z}_i$ is chosen such that its minimum distance to all points in $\mathbf{Z}_i$ is largest, namely $\mathbf{z}_{i+1} = \text{argmax}_{\mathbf{z} \in \mathbf{Z}/\mathbf{Z}_i} \left(\min_{\hat{\mathbf{z}} \in \mathbf{Z}_i} d(\mathbf{z}, \hat{\mathbf{z}}) \right)$, where $d(\cdot,\cdot)$ is cosine distance, a common distance measure for comparing word embeddings [30, 36]. This provides diversity in the sampling of documents. Next, for each document we select its representative word as one closest to the document, measured via d, under several constraints. If the closest word has already been sampled before, or its string edit distance to an already sampled word is within a threshold, set to 0.8 in our experiments, then we discard it and find the next closest word. We repeat this process until a word satisfies these conditions. We also restrict words to have a corpus frequency count of a certain threshold, in our case set to 500, so that the sampled words are meaningful to the user. We denote the resulting sampled word vectors as the sequence $\mathbf{W}_s$ for s sampled words. Sampled in this way, words are both semantically distinct and represent the diversity in the documents.

The next step is to perform a 2D projection of the sampled words. We use the Barnes-Hut t-SNE scheme [40] for this purpose, which applied to $\mathbf{W}_s$ gives us $\bar{\mathbf{W}}_s = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_s) \subset \mathbb{R}^2$. Figure 6 shows an example for 100 words, where each word is color and opacity-mapped based on its index: purple and opaque words are sampled early, while green and transparent words are sampled late in the sequence. Note how words are well-spaced with respect to their order in the sequence. This highlights t-SNE's capabilities of pushing dissimilar objects far away in the projection space, better discrimination of similar words, and ultimately satisfying our multi-scale sampling criterion.

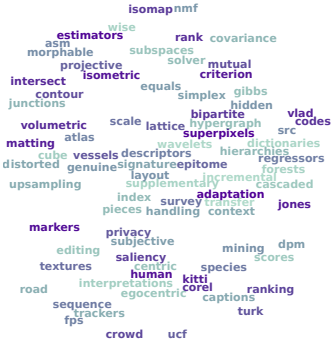


Fig. 6. 2D projection of sampled words.

4.5 Document Projections

In order to dynamically project documents that adhere to user interactions, we treat the sampled words $\mathbf{W}_s$ as a means for a user to specify concepts. The linear structure in the embedding permits us to represent a document as a summation of word vectors, as well as document vectors. We utilize this by having each word $\mathbf{w} \in \mathbf{W}_s$ take on the meaning of a set of user-defined concepts, and project documents based on these concepts. We represent each concept as a vector in the embedding space, and denote the resulting set of vectors as $\mathbf{C}$. Each vector encodes three basic concepts: arbitrary phrases, documents, or document-phrase analogies – we defer discussion on the specific ways user prescribe concepts to the next section on cite2vec's interface. Given the concept set $\mathbf{C}$, we interpret each word as taking on the meaning of itself composed with any concept in $\mathbf{C}$: $\mathbf{w} \in \mathbf{W}_s$ can be $\mathbf{w}+\mathbf{c}$ for $\mathbf{c} \in \mathbf{C}'$.

In projecting document vectors, we ensure that the word projections $\bar{\mathbf{W}}_s$ remain fixed, and the document projections satisfy the following:

- Each document is projected close to its best matching word-concept composition – the concepts effectively steer the document projection.
- The geometric structure of the embedding space of the documents should be preserved as much as possible. This attempts to preserve document-document relationships.

There are two key challenges in constructing such a projection. First, straightforwardly applying a dimensionality reduction approach like t-SNE with the fixed 2D word projections as hard constraints results in non-metric composition-document constraints, since a single word may have different concept compositions, depending on the document. Secondly, we cannot exclusively rely on words to retain document-document relationships in the projection, since the set of best matching word-concept compositions may correspond to words that are far apart in the 2D projection space. This is due to a document being potentially represented by a diverse set of compositions.

To satisfy the first desideratum, we first find document $\mathbf{z}_i$'s 2D word projection $\mathbf{p}_{i'}$ whose corresponding embedding vector $\mathbf{w}_{i'}$ best composes with $\mathbf{C}$, namely:

$$\mathbf{w}_{i'} = \underset{\mathbf{w} \in \mathbf{W}_s}{\operatorname{argmin}} \left(\min_{\mathbf{c} \in \mathbf{C}} d(\mathbf{z}_i, \mathbf{w}+\mathbf{c}) \right), \tag{3}$$

and ensure that $\mathbf{z}_i$'s 2D coordinate $\mathbf{q}_i$ is close to $\mathbf{p}_{i'}$. This enforces each document to be close to its best-matching concept composition – notice that different documents can be assigned to the same word, but composed with different concepts. To satisfy the geometric criterion, we employ regularization with respect to the graph Laplacian of the document embedding – a common form of regularization in preserving geometric structure [44], enforcing each document to be a weighted average of similar documents. This is similar to the Least Squares Projection approach [33] and related techniques [34], however a major difference is in the choice of control points: our joint embedding permits us to select semantically meaningful 2D positions with respect to the user-defined concepts.

More specifically, consider matrix $\mathbf{P}' \in \mathbb{R}^{n_d x 2}$ as the 2D projections corresponding to each document's nearest neighbor word, i.e. the i'th row of $\mathbf{P}'$ is $\mathbf{p}_{i'}$. We denote the document graph Laplacian as Δ_d, constructed through a k nearest neighbor-based affinity matrix between document vectors, where we have set $k = 7$. The cosine distance $d(\cdot,\cdot)$ is used as the distance measure, taken as the argument of a Gaussian kernel to form the affinities. The documents' 2D projections $\mathbf{Q} \in \mathbb{R}^{n_d x 2}$ are found by prescribing $\mathbf{P}'$ as soft constraints, while minimizing the document Laplacian with respect to $\mathbf{Q}$, leading to the following linear system:

$$(\Delta_d + \alpha \mathbf{I})\mathbf{Q} = \alpha \mathbf{P}', \tag{4}$$

where α controls the influence of the word 2D projections, set to 10 in our experiments. Intuitively, if each document's projection is explicitly assigned its closest word embedding's 2D projection, then it is possible for many documents to map to the same location. By enforcing a Laplacian regularization, these points then disperse in the 2D space in order to reflect the intrinsic geometric structure of the document embedding.

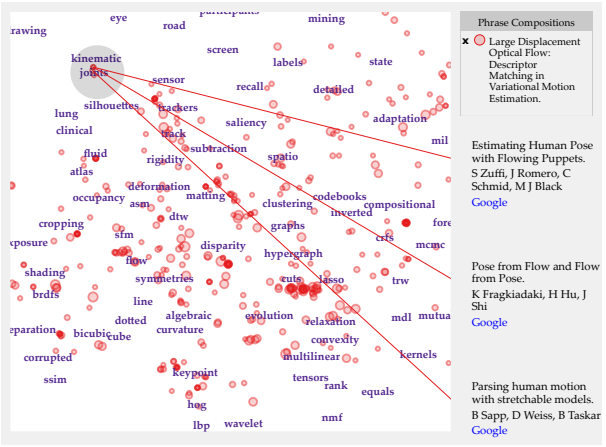


Fig. 7. Documents-as-concepts: the user prescribes all projected words to be composed with the selected document (shown in the upper right), and all documents are positioned to satisfy the concept composition.

5 Exploring Word-Document Embeddings

Equipped with a mechanism for projecting words and documents, we now discuss our visualization interface for enabling the user to search documents via their usage.

Word Exploration. We would like the user to interact with the word projection in order to inspire further search as well as understand word/document similarities through their spatialization. A challenge in presenting words, however, is clutter. Note that the words are sampled in a multi-scale manner, such that concepts become progressively more specific the further in the sampled word sequence. We exploit this property by enabling the user to pan and zoom in the 2D interface, such that the number of words shown is a function of the zoom level. Although very straightforward, this reduces clutter while gradually showing more specific concepts as the user zooms, though more sophisticated word layout techniques could be used [35].

Document Visualization. Given the words, our interface initially shows a projection of the documents absent of any concepts, i.e. $\mathbf{C} = \emptyset$ and documents are positioned near their best-matching words via Equation 4. We represent each document's 2D position with a disk, sized proportional to the distance to its corresponding word, and assign each disk a certain amount of opacity so that clusters of documents become apparent via the accumulation of the disks' opacities. We also provide a document lens that shows information on those documents whose 2D coordinates are inside of the lens' disk, akin to excentric labeling [18] and CiteRivers [21]. We also provide a Google link to search for the document, allowing the user to read the document and better understand its placement in the projection space. For our computer vision dataset, document citations are not always research papers, but can be project websites hosting code, data, or benchmarks – our interface allows the user to quickly discover these resources by searching the projection space.

Specifying a Concept. The above projection positions documents based on their closest displayed word. However, as documents may be used in many different ways, we would like the user to explore document usage via the semantics in which other people cite documents. We achieve this by allowing the user to specify an arbitrary phrase as a concept, and assign a single vector to $\mathbf{C}$ as the summation of the phrases' word vectors, i.e. each projected word is now composed with the phrase.

Given a phrase we first solve for the documents' positions via Equation 4. We then filter out documents if their closest concept composition has a large cosine distance – a threshold we set to 1 in all experiments.

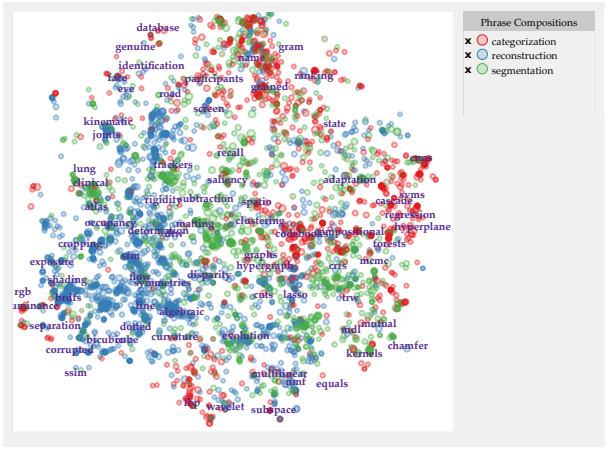


Fig. 8. Comparing concepts: the user specifies three coarse-level concepts, enabling them to observe distinct themes based on the document distribution and document assignment to concepts, while also discovering words that share concepts.

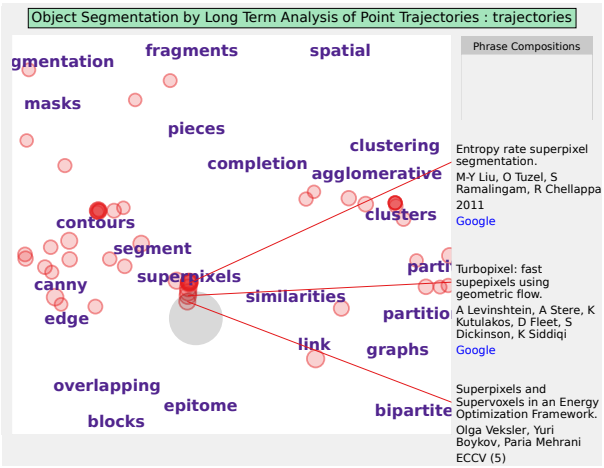


Fig. 10. Analogy concepts: the user may specify the first part of a document:word analogy (shown on top), resulting in a new document projection such that a document's proximity to a word indicates their likely completion of the analogy.

This conveys to the user what concept compositions are appropriate for the documents. We next animate the retained documents' disks to move from their current positions to the newly solved positions. In practice, the retained documents tend to be initially clustered in a single region of the space, but the animation will reveal a dispersion to a diverse set of concepts due to better matching word-concept compositions – see Figure 2. This allows the user to see the variety of ways in which documents are used.

We also allow the user to specify documents themselves as concepts, wherein each projected word is now composed with a document. We support this via the document lens: we allow the user to anchor the lens, and select the document title as a concept, assigning $\mathbf{C}$ as the document's vector. Documents embody a very rich set of meanings, and in practice we find this to be a very useful way of specifying concept compositions if the user is familiar with a document. Figure 7 shows an example where the specified paper uses multiple concepts: optical flow, motion, variational methods, and the projection yields papers where these concepts are used in conjunction with human pose, as indicated by the "kinematics" and "joints" words. Hence documents-as-concepts enables the user to explore usage compositions with diverse meanings.

Comparing Concepts. In our interface we also allow the user to specify multiple concepts – either phrases or documents – in order to enable the user to compare document usage. More specifically, upon specification of a new concept, we add it to $\mathbf{C}$, in addition to all *pairwise combinations* of concepts. This is done to enrich the space of concepts, so that the user can discover finer-grained compositions that characterize documents.

Given the new set $\mathbf{C}$ we again solve the Laplacian system of Equation 4. In our interface we strive to maintain continuity as a user adds concepts. We achieve this by first filtering and animating the retained documents as before, and then fade in documents that meet the distance threshold for the newly added concept. Furthermore, we assign a unique color to each concept, and color each document based on its best-matching concept, so that the user can distinguish documents of different concept compositions. For pairwise combinations, we assign the fill of a disk the color of one concept, and its stroke the other concept's color.

Figure 8 shows an example of comparing document usage via different concepts. The three concepts – categorization, reconstruction, and segmentation – are dominant topics in computer vision, and as shown documents are positioned and matched with concepts that are fairly distinct with respect to the projected words. We can see, however, that all three concepts lie near "graphs", and indeed it is quite common for approaches in these three areas to use graphs as part of their approach. Hence we see how concept comparison enables the user to discover distinct regions of usage, as well as usage overlap, in exploring documents.

Exploring Analogies. We also enable analogies to be specified as concepts. In our interface, the user specifies the first part of an analogy in the form of a document:word pair. We form a concept vector from the analogy as the difference between the word and document, assign to $\mathbf{C}$ this vector, and find new document positions via Equation 4, similarly filtering and animating documents as discussed above. The projected words now take on the concept of *analogy completions*: a document's proximity to a word indicates the completion of the second half of the analogy.

As analogies may not always be obvious for a user to specify, we have devised a simple method of analogy suggestion. Upon selecting a document, we find a small number of representative words, such that their nonnegative linear combination of vectors is close to the document vector. We find these words and their weights via nonnegative sparse coding [39], which will produce a sparse representation of a document via its words, often finding rather diverse concepts. In our interface, we show the representative words, and their weights, as a tooltip below the selected document, as shown in Figure 9. From this, the user can select a word to form the analogy concept.

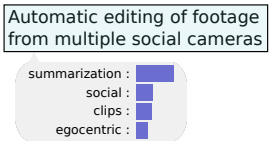
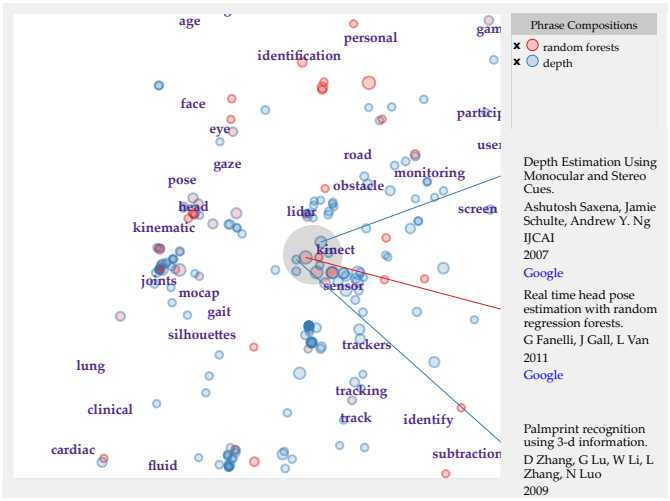


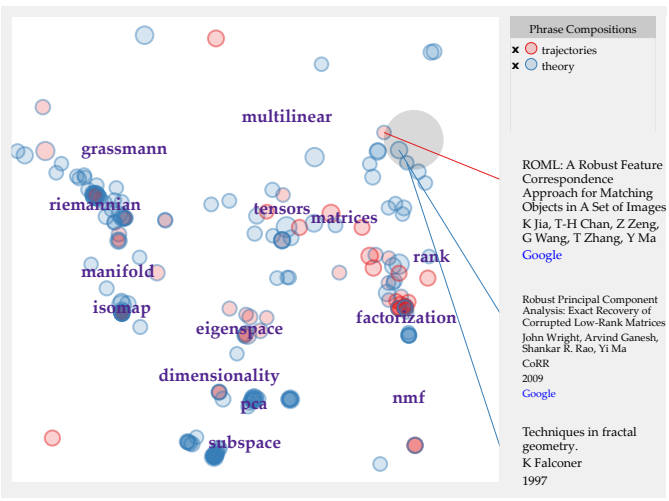
Fig. 9. Example analogy suggestion.

Figure 10 shows an example of analogical exploration. The user-supplied analogy is shown on top, and in the projection space, each paper's proximity to a word indicates its likelihood of completing the analogy. In this example, the supplied analogy conveys approaches focused on object segmentation that use trajectory data, hence the underlying analogy is: segmentation approaches which use a certain type of data. Analogy completions lead to the discovery of object segmentation approaches which use superpixels as data, centered at the superpixel word. We observe that analogies can explain document-word relationships by example, leading to richer discovery of concepts in the document collection.

Enabling Interactivity. All of the above operations rely on an interactive means of updating document projections. Note that in solving for Equation 4 we only need to update each paper's closest composed word, i.e. only the right hand side of the equation. Hence as a preprocess, we factorize $(\Delta_d + \alpha \mathbf{I})$, so that at runtime we can quickly perform back-substitution with the new $\mathbf{P}'$. This results in low latency as a user edits concepts, at most a couple seconds even for a modest number of concepts, i.e. 4-5.



(a) Phrase Compositions: Depth-Based Random Forest Techniques



(b) Phrase Compositions: Theoretical Trajectory Techniques

Fig. 11. We show use cases for exploring document usage via multiple phrase compositions. In (a) we highlight how standard domain-specific phrases lead to basic document usage, namely techniques and data that are used by such documents. In (b) we highlight how the phrase "theory" leads to papers that are naturally theoretically-based, a concept that would be difficult to discover via the papers themselves, but becomes apparent in observing citation contexts.

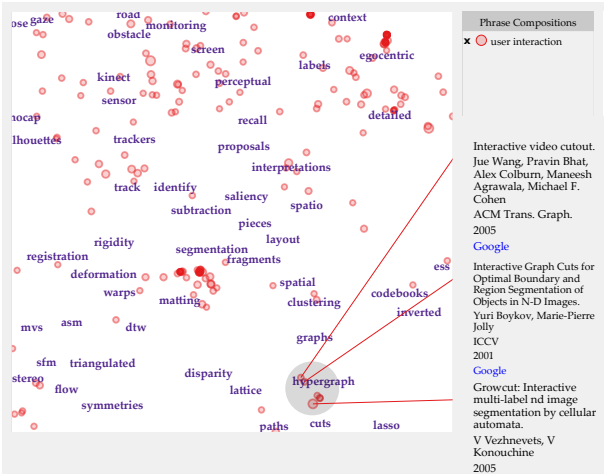


Fig. 12. We show how a non-technical phrase, user interaction, leads to different techniques that use interaction as part of their approach, here highlighting interactive segmentation techniques.

6 Experiments

We first highlight the various aspects of cite2vec through some simple use-cases. We then perform two separate case studies: a detailed study in which a person is researching an unfamiliar research area, and a broader study in which we allow participants of a computer vision conference to demo our system.

6.1 Searching Computer Vision

In Figures 11(a) and (b) we show the benefits of specifying multiple phrases for document exploration. In Figure 11(a) we seek approaches which use the technique of random forests, as well as those approaches that process depth-based data. We highlight a case where the combination of these phrases composes with "kinect" in the word projection, leading to a paper focused on pose estimation from single view-depth acquisition devices (i.e. such as the Microsoft Kinect) using random forests, shown as the disk filled in red and outlined in blue. In Figure 11(b) we seek approaches which process trajectories, as well as approaches deemed theoretical. As shown, we discover a paper ("ROML" in the figure) that is focused on trajectory-based data – image sequences – but is also using concepts from low-rank modeling and providing theoretical bounds on their approach. Concepts such as "theory" are difficult to represent in document-based approaches, since the language of a theory paper is fairly technical and distinct. Our citation-driven approach, however, can easily tease this out since it is common for a theory paper to be cited as such by other research papers, as people tend to be specific on the nature of a citing paper's research contributions.

We can also search documents via non-technical phrases, where in Figure 12 we wish to see if people cite documents because they have an element of user interaction. We observe that the resulting projection leads to a number of interactive techniques focused on segmentation. Indeed, we end up finding a number of segmentation methods that employ user interaction, in particular methods which employ graph-based techniques, as evident by the surrounding words.

6.2 Case Study: Researching Weak Supervision

In this case study we have a fellow researcher, whose expertise is in machine learning, use our system to research an unfamiliar subdiscipline: weakly supervised learning techniques in computer vision. The user is interested in finding related research to weak supervision methods, as well as more detailed techniques to weak supervision which use deep learning methods.

Upon searching for the phrase "weak supervision", the user found several interesting patterns. First, several distinct document clusters formed over particular word sets: domain adaptation, attributes, and boosting techniques. The researcher confirmed that each of these concepts are highly related to weak supervision, which highlights the fact that cite2vec is capable of grouping similar concepts based on the context in which these documents are cited. In this case, citations pertaining to weak supervision co-occur with these other subdisciplines.

Upon finding the attributes word and the collection of papers from the weak supervision search, the user then decided to do a new search with "attributes" as one phrase, as well as "convolutional neural networks" as another phrase, as he was interested to see if there existed any prior work which used these concepts. Under the word for aesthetics, the user found a paper devoted to learning clothing style from attributes and convolutional neural networks. The user was unsure if any work had been done which combined the two, but was pleased to find this particular paper, placed in a semantically meaningful portion of the space, i.e. aesthetics composed with attributes and convolutional

neural networks. This demonstrates the capability of our method in discovering complex compositions of concepts.

6.3 Case Study: General Usage

In our second case study, we attended the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) as an exhibitor in order to demonstrate our system. Our goal was to elicit more general feedback from a broad audience, where the attendees all have a certain level of expertise in some area of computer vision. We asked each participant to input a phrase of choice into our system, and explore the resulting document projections tailored to their research interests. At the time of attendance, our system only provided support for word compositions; nevertheless, we feel that the feedback is reflective of our general approach. In total, we had 21 attendees demo our system, each spending on average about 5 minutes with our system.

Overall we received encouraging feedback from the conference attendees. Several attendees acknowledged the challenges in keeping up with the rapid pace of computer vision publications, recognizing the need for more advanced systems that are easier to use than current technologies, for instance Google Scholar, are necessary to search for papers, and found our system to be a better means of searching than such conventional techniques. One attendee remarked that traditional means of searching for papers, in particular via citation count statistics, can potentially bias the research direction of the community, and found our system to be a potential alternative.

The attendees used a wide variety of phrases to explore the document space. Some were more traditional computer vision topics, such as key pose, fine grained classification, photometric stereo, and visual tracking. In these cases, attendees found our composition interface comprehensible, and remarked that the resulting paper projections generally made sense with respect to their associated word-phrase compositions, with one attendee mentioning that the system could be used to "construct references for a paper." Other attendees used more unconventional or specific phrases: iris, transportation, hardware. Yet the resulting composed document projections made sense to the attendees: in these cases returning results on biometric systems, outdoor scenes and roads, and documents related to code and efficiency, respectively, with the iris search prompting the attendee to state that the system was "very useful". In particular, the user who searched for hardware found a number of documents centered around the deep learning portion of the projection space, and was happy to see various software packages that support deep learning methods (i.e. Caffe, Theano, Torch, etc..).

Interestingly, several attendees searched for the last name of authors in our system. Note that in our citation-based approach, authors will only appear if they are explicitly mentioned in the text, whether or not they are mentioned near specific citations. However, in all cases, the resulting projections were intuitive to the attendees: not only did papers appear whose author lists contained the author, but other papers showed whose research areas overlap with the author's interests. We found these results to be very encouraging, suggesting the quality of the learned embeddings and effectiveness of our interface.

7 Discussion

Our experiments showcase the effectiveness of our system and its potential efficacy. However, we see a number of limitations with our approach, many of which are motivated by our case studies.

7.1 Inclusion of Meta-data

As observed by the researcher in the detailed case study, one limitation of our approach is that it is not particularly effective for finding so-called "hubs" in certain subdisciplines, e.g. well-established and prominent documents. Although the researcher found cite2vec to be very effective for general search, he found the interface somewhat overwhelming if one wants to find papers deemed important. Measures of importance are frequently derived through meta-data in document collections, e.g. in research papers one can use citation counts as a basic statistic. Supporting such a feature, however, is ultimately user-dependent; as mentioned in the WACV case study, one attendee found a potential risk in prioritizing search based on this. Nevertheless, we think that it should be possible to incorporate such an importance-based feature as an option for our system.

In the WACV case study, we received much feedback on the incorporation of other forms of meta-data into our approach. These consist of authors, publication venues and publication years. We think that our base system can be made into a much richer visual analytics tool by incorporating such meta-data, allowing for the user to filter papers based on these fields, in conjunction with our proposed method for searching. Alternatively, another option is to include such meta-data into the embedding itself, and learn vector representations of authors, venues, even years, based on citation contexts. This could provide for a rich semantic meaning of such concepts.

One WACV attendee observed that it would be nice to have the system adapt to the user behavior by *learning* from user interactions. We find such an approach very promising, in particular for the correction of any false positives found by the user. We think that recent word embedding approaches which learn from side information [29] can be used in conjunction with user feedback, to interactively build better neural language models for document collections.

7.2 Data Requirements and Scalability

A limitation of our approach is the dependency on a sufficient number of citation contexts in modeling documents – e.g. a newly published paper is unlikely to have sufficient citation evidence to be incorporated into the embedding. It could be interesting to combine traditional topic models, like LDA, with our technique to address the citation-scarcity issue. On the other side of scalability, although our approach scales well to 20,000-30,000 documents, past this our approach has issues in maintaining interactivity in dynamic document projections, especially as the number of user concepts increases. However, we think hashing-based schemes [12] can help in finding approximate nearest neighbor word-concept compositions in an efficient manner.

7.3 Embeddings for Visualization Tasks

We think that word embeddings, in a more general sense, have a role in other areas of visualization - not even necessarily just text. One can view word2vec as an unsupervised learning scheme for *predicting context*. Recent approaches in computer vision [14, 26] have capitalized on this for learning good visual representations, using either spatial context or visually grounded context as supervisory signal. It should be possible to apply these ideas to various avenues of visualization where some form of context can be defined, such as graph layouts [25] or visual palettes [13], leading to useful data representations that provide effective ways of interacting with data.

8 Conclusion

We have presented a novel method for modeling and exploring documents via their citation contexts. Our approach, cite2vec, permits the user to explore citation-based meanings of documents, allowing the user to express word/document compositions and document:word analogies in a 2D projection space, tailoring the projection of documents to the user's interests. This is enabled via learning an embedding space in which both words and documents exist, where proximity of words and/or documents reflects their underlying semantic similarity, determined via word/document co-occurrences in local context windows of unstructured text. Our experiments demonstrate the promise of our approach for document exploration.

For future work, we would like to incorporate time into our model in order to understand how word/document semantics change as a function of time. As citation contexts play the essential role in document representations, visualizing the contexts, as well as prioritizing contexts, would be useful to incorporate, leveraging tools such as Serendip [1]. We also intend to incorporate the feedback obtained via our case studies in order to build a visual analytics tool that supports a diverse range of functionality in exploring document collections.

9 Acknowledgements

We thank the reviewers for their helpful suggestions in improving the paper.

REFERENCES

[1] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher. Serendip: Topic model-driven visual exploration of text corpora. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pages 173–182. IEEE, 2014.

[2] E. Amorim, E. V. Brazil, L. G. Nonato, F. Samavati, and M. C. Sousa. Multidimensional projection with radial basis function and control points selection. In *2014 IEEE Pacific Visualization Symposium*, pages 209–216. IEEE, 2014.

[3] F. Beck, S. Koch, and D. Weiskopf. Visual analysis and dissemination of scientific literature collections with survis. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):180–189, 2016.

[4] C. T. Bergstrom, J. D. West, and M. A. Wiseman. The eigenfactor metrics. *The Journal of Neuroscience*, 28(45):11433–11434, 2008.

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

[6] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2271–2280, 2014.

[7] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*, pages 288–296, 2009.

[8] C. Chen. Citespace ii: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for information Science and Technology*, 57(3):359–377, 2006.

[9] J. Choo, C. Lee, C. K. Reddy, and H. Park. Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):1992–2001, 2013.

[10] I. G. Councill, C. L. Giles, and M.-Y. Kan. Parscit: an open-source crf reference string parsing package. In *LREC*, 2008.

[11] P. J. Crossno, D. M. Dunlavy, and T. M. Shead. Lsaview: a tool for visual exploration of latent semantic modeling. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 83–90. IEEE, 2009.

[12] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 253–262. ACM, 2004.

[13] Ç. Demiralp, M. S. Bernstein, and J. Heer. Learning perceptual kernels for visualization design. *IEEE transactions on visualization and computer graphics*, 20(12):1933–1942, 2014.

[14] C. Doersch, A. Gupta, and A. A. Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1422–1430, 2015.

[15] W. Dou, X. Wang, R. Chang, and W. Ribarsky. Paralleltopics: A probabilistic approach to exploring document collections. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 231–240. IEEE, 2011.

[16] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2002–2011, 2013.

[17] C. Dunne, B. Shneiderman, R. Gove, J. Klavans, and B. Dorr. Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. *Journal of the American Society for Information Science and Technology*, 63(12):2351–2369, 2012.

[18] J.-D. Fekete and C. Plaisant. Excentric labeling: dynamic neighborhood labeling for data visualization. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 512–519. ACM, 1999.

[19] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1646–1663, 2013.

[20] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):9–20, 2002.

[21] F. Heimerl, Q. Han, and S. Koch. Citerivers: visual analytics of citation patterns. *Visualization and Computer Graphics, IEEE Transactions on*, 22(1):190–199, 2016.

[22] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. L. Giles. A neural probabilistic model for context based citation recommendation. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[23] S. Ingram and T. Munzner. Dimensionality reduction for documents with nearest neighbor queries. *Neurocomputing*, 150:557–569, 2015.

[24] P. Joia, D. Coimbra, J. A. Cuminato, F. V. Paulovich, and L. G. Nonato. Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2563–2571, 2011.

[25] S. Kieffer, T. Dwyer, K. Marriott, and M. Wybrow. Hola: Human-like orthogonal network layout. *IEEE transactions on visualization and computer graphics*, 22(1):349–358, 2016.

[26] S. Kottur, R. Vedantam, J. M. F. Moura, and D. Parikh. Visual word2vec (vis-w2v): Learning visually grounded word embeddings using abstract scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[27] H. Lee, J. Kihm, J. Choo, J. Stasko, and H. Park. ivisclustering: An interactive visual document clustering via topic modeling. In *Computer Graphics Forum*, volume 31, pages 1155–1164. Wiley Online Library, 2012.

[28] O. Levy and Y. Goldberg. Dependency-based word embeddings. In *ACL (2)*, pages 302–308, 2014.

[29] Q. Liu, H. Jiang, S. Wei, Z.-H. Ling, and Y. Hu. Learning semantic word embeddings based on ordinal knowledge constraints. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 1501–1511, 2015.

[30] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[31] D. Oelke, H. Strobelt, C. Rohrdantz, I. Gurevych, and O. Deussen. Comparative exploration of document collections: a visual analytics approach. In *Computer Graphics Forum*, volume 33, pages 201–210. Wiley Online Library, 2014.

[32] F. V. Paulovich and R. Minghim. Hipp: A novel hierarchical point placement strategy and its application to the exploration of document collections. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1229–1236, 2008.

[33] F. V. Paulovich, L. G. Nonato, R. Minghim, and H. Levkowitz. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *Visualization and Computer Graphics, IEEE Transactions on*, 14(3):564–575, 2008.

[34] F. V. Paulovich, C. T. Silva, and L. G. Nonato. Two-phase mapping for projecting massive data sets. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1281–1290, 2010.

[35] F. V. Paulovich, F. Toledo, G. P. Telles, R. Minghim, and L. G. Nonato. Semantic wordification of document collections. In *Computer Graphics Forum*, volume 31, pages 1145–1153. Wiley Online Library, 2012.

[36] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.

[37] D. Shahaf, C. Guestrin, and E. Horvitz. Metro maps of science. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1122–1130. ACM, 2012.

[38] A. Skupin. A cartographic approach to visualizing conference abstracts. *Computer Graphics and Applications, IEEE*, 22(1):50–58, 2002.

[39] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[40] L. Van Der Maaten. Accelerating t-sne using tree-based algorithms. *The Journal of Machine Learning Research*, 15(1):3221–3245, 2014.

[41] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. Tiara: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 153–162. ACM, 2010.

[42] J. A. Wise. The ecological approach to text visualization. *Journal of the Association for Information Science and Technology*, 50(13):1224, 1999.

[43] J. Zhang, C. Chen, and J. Li. Visualizing the intellectual structure with paper-reference matrices. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1153–1160, 2009.

[44] X. Zhu, Z. Ghahramani, J. Lafferty, et al. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, volume 3, pages 912–919, 2003.