

# Evaluating Semantic Interaction on Word Embeddings via Simulation

Yali Bian\*  
Virginia Tech

Michelle Dowling†  
Virginia Tech

Chris North‡  
Virginia Tech

## ABSTRACT

Semantic interaction (*SI*) attempts to learn the user’s cognitive intents as they directly manipulate data projections during sensemaking activity. For text analysis, prior implementations of *SI* have used common data features, such as bag-of-words representations, for machine learning from user interactions. Instead, we hypothesize that features derived from deep learning word embeddings will enable *SI* to better capture the user’s subtle intents. However, evaluating these effects is difficult. *SI* systems are usually evaluated by a human-centred qualitative approach, by observing the utility and effectiveness of the application for end-users. This approach has drawbacks in terms of replicability, scalability, and objectiveness, which makes it hard to perform convincing contrast experiments between different *SI* models. To tackle this problem, we explore a quantitative algorithm-centered analysis as a complementary evaluation approach, by simulating users’ interactions and calculating the accuracy of the learned model. We use these methods to compare word-embeddings to bag-of-words features for *SI*.

**Keywords:** Semantic interaction, evaluation, word embedding, sensemaking, visual text analytics, deep neural embedding

## 1 INTRODUCTION

Semantic interaction (*SI*) [10, 11] is an interaction technique for non-experts to interact with the underlying algorithms of visual analytics (*VA*) systems [7]. With *SI*, analysts can focus on reasoning and manipulating data in the 2D spatialization instead of directly interacting with the underlying machine learning (*ML*) models [1]. It is, therefore, the system’s responsibility to tune the *ML* models by capturing users’ interactions and inferring the analyst’s intent [25]. These intents are then translated to updated model parameters via semi-supervised machine learning techniques (*ML*) [10], such as metric learning [5]. Hence, analysts can remain concentrated on their sensemaking activities [23].

In this paper, we investigate two questions about *SI*: (1) Can word embeddings help *SI* learn user’s interactive intents better than traditional features? (2) How can we comparatively evaluate alternative *SI* models, such as in question (1)?

The human-centred approach [3] is the primary evaluation method for *SI* systems. The *SI* system is measured by the utility and effectiveness of the application for end-users through studies of human subjects. For example, ForceSPIRE [12] is a visual text analysis system powered by *SI*, and its effectiveness was measured by analysts’ performance on an intelligence analysis task. This approach is highly dependent on human feedback, which leads to several challenges when comparing the effectiveness of different *SI* models because: user interactions are difficult to precisely replicate; *SI* interactions are incremental and iteratively build; does not scale well to compare many model alternatives as needed in *ML*; and usage differences can mask subtle model performance differences.

\*e-mail: yali@vt.edu

†e-mail: dowlingm@vt.edu

‡e-mail: north@vt.edu

To cope with these problems, we design and perform an additional quantitative algorithm-centred analysis that simulates human interactions as a complement to human-centered evaluation. In the evaluation, we use the labelled text datasets as the ground truth, since there is no ground truth for the users complex intents and concepts. Then the ground truth is used to simulate user’s interactions, and we calculate and compare the accuracy of trained models.

In this paper, we first describe two alternative *SI* models for visual text analysis using different data features as inputs (embedding vectors, and keywords vectors). We then implement a generalized visual text analysis prototype with *SI* that can take both embedding vectors and keywords vectors as features. We then demonstrate how the two kinds of evaluation methods together provide a complementary evaluation and comparison of these two *SI* models.

## 2 SEMANTIC INTERACTION WITH WORD EMBEDDINGS

We investigate the opportunity of using word embeddings to better support *SI*. *SI*’s ability to infer the analyst’s reasoning process is fundamentally limited by the feature space of the underlying machine learning. Text analytics systems with *SI*, such as ForceSPIRE [4, 12] and Cosmos [9], typically use keywords (such as text terms and phrases) as data features ( $SI_{keyword}$  in Fig. 1a), known as bag-of-words [31]. Recently, word embedding techniques [6, 20, 22, 30], also known as deep learning representations [2, 18, 26, 29], have shown significant advantages in numerous tasks in natural language processing and information retrieval [15, 17, 28] over bag-of-words features. Therefore, the combination of *SI* with word embedding ( $SI_{embedding}$  in Fig. 1) might enable better learning ability than  $SI_{keyword}$  to update the underlying machine learning models [19]. The intuition is that word embeddings could represent more abstract concepts that are closer to modeling human cognitive reasoning. To test the hypothesis, we focus on the comparisons of  $SI_{embedding}$  with  $SI_{keyword}$  through the same *SI* pipeline but with different document features as input:

- **Keyword features used in  $SI_{keyword}$ :** We use TF-IDF values as the keyword features, and word hashing to compress the large number of words from the document collection into 300 dimensions.
- **Embedding features used in  $SI_{embedding}$ :** We use the pre-trained GloVe model [21] with 300 dimensions to extract embedding features, and use the “basic averaged word embeddings” method to compute average word embedding for all words in a document [17].

### 2.1 Application Prototyping

As shown in Fig. 1,  $SI_{embedding}$  has a similar structure with  $SI_{keyword}$  as they both use numerical vectors to represent documents (embedding vectors, and TF-IDF vectors). We are able to switch between  $SI_{keyword}$  and  $SI_{embedding}$  in the same *SI* prototype. We build the prototype upon the foundation of Andromeda [24], a visual analytics tool for exploring high-dimensional data projections. The prototype can use either bag-of-words or embedding vectors as features and update the feature weights to capture analysts’ intents. For bag-of-words, the system will up-weight the shared words in the dragged documents in response to the analyst dragging two or more

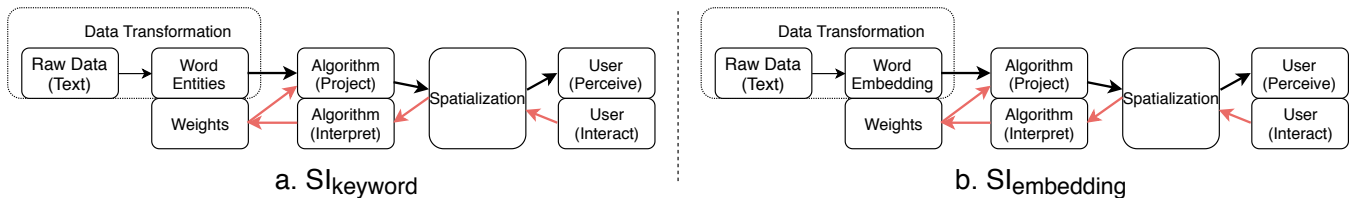


Figure 1: In the SI pipeline, distance metric learning interprets users’ interactions on the projection. (a) In  $SI_{keyword}$ , the extracted features of text data are keywords; (b) In  $SI_{embedding}$ , the features are embedding vectors.

documents closer together. For embedding vectors, the system will up-weight the dimensions of the embedding vectors with similar patterns in the dragged documents.

As shown in Figure 2-1, the visualization is a projection of documents. The distance between documents reflects their relative similarity according to a weighted distance metric over their features. At first, documents scatter in the workspace based on their features and an equally-weighted dimension reduction of the high-dimensional representation. If two documents are positioned close to each other in this initial projection, it implies that these two documents are similar based on all vector features. Semantic interaction enables analysts to directly manipulate the projection by moving the documents to express their own domain knowledge about desired similarities. With this interaction, analysts can express the semantic relationships between documents, which thereby informs the underlying weighted distance model and updates the projection. Through this, the projection can be customized according to the learned intent of the analyst.

### 3 EVALUATING SI EMBEDDING

Traditionally, the human-centred approach is a good method to test if the SI model is an effective application for analysts to perform sensemaking tasks. However, comparing two SI models ( $SI_{keyword}$  and  $SI_{embedding}$ ) through this approach is inadequate. Thus, we also design a replicable quantitative study to validate  $SI_{embedding}$  in comparison to  $SI_{keyword}$  in 4 simulated text analysis tasks of different difficulty levels.

#### 3.1 User-Centered Qualitative Analysis

In this section, we present a practical use case about intelligence analysis as our qualitative study to compare  $SI_{keyword}$  against  $SI_{embedding}$ . For this analysis, we engaged an expert in intelligence analysis to review the visual result generated from both  $SI_{embedding}$  and  $SI_{keyword}$  to provide qualitative feedback, and provide grounding for the quantitative analysis.

##### 3.1.1 Crescent Dataset

The crescent dataset [13] has 42 fictional intelligence reports regarding a coordinated terrorist plot in Boston, New York, and Atlanta. Only 24 reports are relevant to these plots. The task for our qualitative analysis is to identify these three terrorist threats by using  $SI_{embedding}$  or  $SI_{keyword}$ .

By moving documents in the visualization according to their perceived similarity, analysts express their reasoning process to the system. The ground truth of this task is as follows, which we use to measure the accuracy of the models: **Boston:** cia7, cia8, cia9, cia10, cia11, fbi1, fbi2, fbi21, se3, se4; **New York:** cia11, fbi1, fbi10, fbi13, fbi16, fbi22, fbi25, se2, se3, se4; **Atlanta:** cia4, cia11, fbi1, fbi7, fbi11, fbi15, fbi17, fbi19, fbi20, fbi24, se4, se5; **Irrelevant:** fbi3, fbi4, fbi5, fbi6, fbi8, fbi9, fbi14, fbi18, fbi23, cia1, cia2, cia3, cia5, cia6, se1.

##### 3.1.2 Case study with $SI_{embedding}$

In this first case study, the analyst performs the task by using the prototype system with embedding as the document features. As shown in Figure 2-1, the analyst updates the layout to reflect the perceived similarities between the documents, grouping three documents about “New York” to the top left region of the projection, indicated by the yellow arrows in Figure 2-2, three documents about “Atlanta” to the bottom left region indicated by blue arrows, and three documents about “Boston” to the bottom right part indicated by red arrows.

After the layout updates, some semantic relationships are revealed (Figure 2-3.a). The left cluster in red circle contains documents about “Boston,” the top right cluster in yellow circle contains documents about “NYSE,” and the bottom right cluster in blue circles contains documents about “Atlanta.” We found that there are semantic mappings between this updated layout and the ground truth; all relevant reports that belong to single plot are well placed into each cluster. Reports about the coordination between two or more plots are between clusters. For example, the report “se3” contains information regarding both the “Boston” and “NYSE” plots, so the document is located between these two clusters. Thus,  $SI_{embedding}$  can capture analysts’ intents and update the model features accordingly. The updated layout then reflects the semantic meanings behind the intents.

##### 3.1.3 Case Study Using $SI_{keyword}$

Mirroring our case study with  $SI_{embedding}$ , the analyst performs similar interactions with the prototype. As shown in Figure 2-3.b, from the updated layout, there are no clear boundaries between different plots. Even after continued interactions, the model is unable to properly differentiate between the three terrorist plots. For example, documents about “Atlanta” and “NYSE” are mixed together. Furthermore, the documents that are pulled out of the big central cluster in panel 4 are primarily the user interacted documents from panel 2. This might indicate that the model is over-fitting based on some specific unimportant keywords.

##### 3.1.4 Expert Review

Besides comparing with the ground truth as we mentioned in Section 3.1, we also asked an expert familiar with the dataset to evaluate the updated layout generated in both  $SI_{embedding}$  and  $SI_{keyword}$  (as shown in Figure 2-3.a and Figure 2-3.b). The expert noted that the  $SI_{embedding}$  layout provides more meaningful information about these plots than  $SI_{keyword}$  layout because the documents from different plots are grouped in different regions. In contrast, the layout generated by  $SI_{keyword}$  makes it difficult to clearly distinguish between these plots. For example, documents in the the  $SI_{embedding}$  layout regarding coordination between the plots are placed between the relevant groups. This is exemplified by how the report “se3” is between the “NYSE” group and the “Boston” group since it discusses how terrorists involved in both plots communicated. In the layout produced by  $SI_{keyword}$ , this document lies at the bottom right of the projection, which does not have such an immediately apparent and relevant semantic meaning.

Additionally, several irrelevant reports were distinguished from the groups in the layout produced by  $SI_{embedding}$ , such as “cia5”,

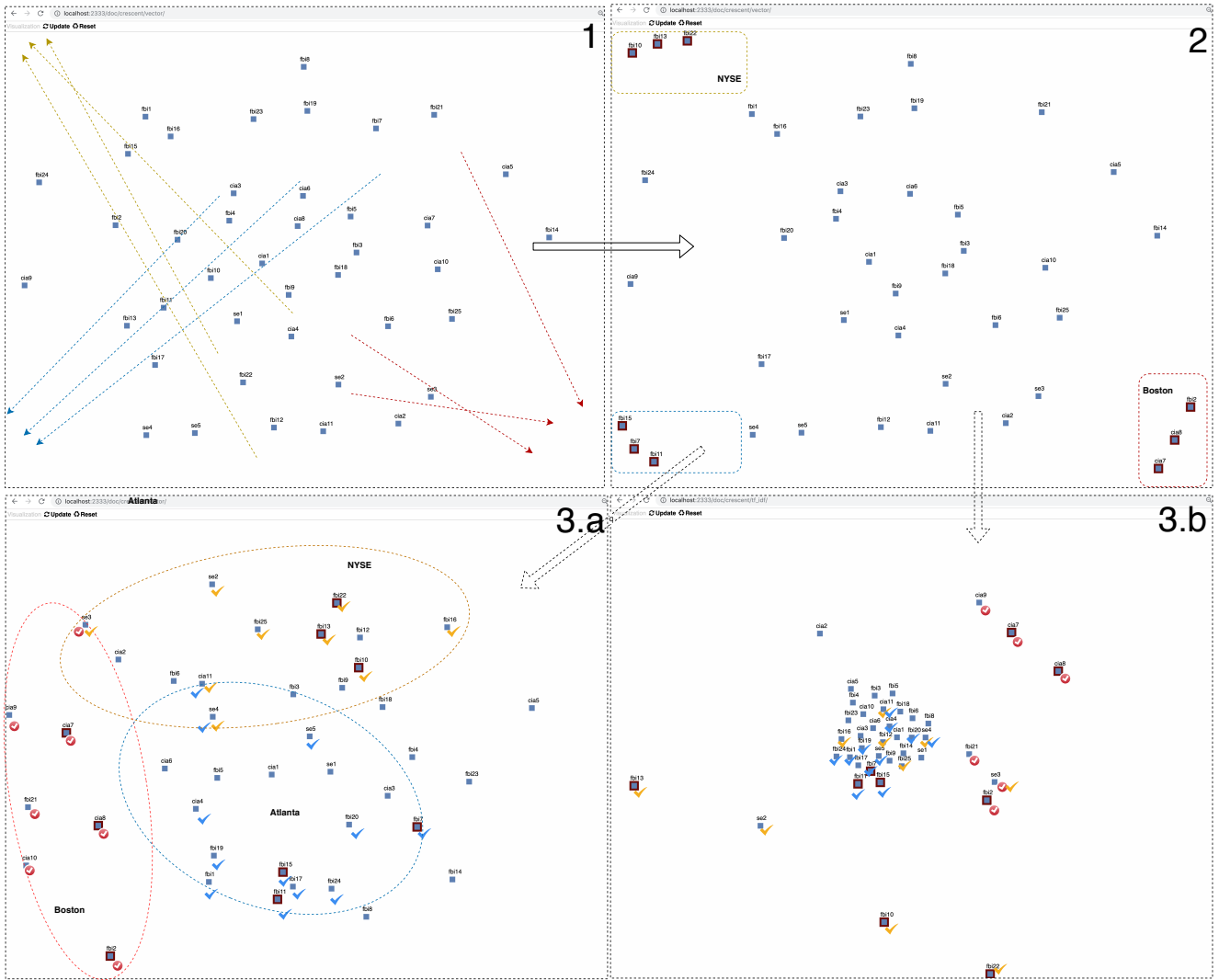


Figure 2: Several screenshots during the two case studies, as discussed in Section 3.1: **Frame 1 and 2** show the similar initial steps performed by analysts in both case studies. **Frame 3.a** shows the resulting projection based on analysts’ interactions, in the case study using  $SI_{embedding}$ . **Frame 3.b** shows the resulting projection based on analysts’ interactions in the case study using  $SI_{keyword}$ .

“fbi23” and “fbi14,” even though they share many keywords with other relevant reports. In contrast, the layout produced by  $SI_{keyword}$  only causes one irrelevant report (“cia2”) to be pulled away from other reports.

### 3.2 Algorithm-centered Quantitative Analysis

In this subsection, we describe our quantitative comparison between  $S_{keyword}$  and  $SI_{embedding}$ . We begin with the experimental design, followed by a description of the datasets we choose for the experiment. Finally, we show the evaluation results and discuss the SI inference abilities of the two models.

#### 3.2.1 Experiment Setup

We need ground truth to evaluate the underlying algorithms used in  $SI_{keyword}$  and  $SI_{embedding}$  quantitatively. There is no ground truth about users’ complex intents and interactions, such as a labeled intelligence hypothesis in a 2D projection. However, there are sufficient text datasets labelled for classification, such as 20newsgroup (section 3.2.2). The labels can be used to simulate different positions in the projection: for example, documents with negative label can be

located at the top left part of the spatial layout, and documents with positive label can be located at the bottom right part. Therefore, we can use classification datasets mapped in the 2D projection as the ground truth, and we only evaluate the inference ability to capture the intents embedded in classification tasks. Then, subsets of the documents from different positions in the 2D projection will be picked and used as simulated semantic interactions from analysts.

Furthermore, since *incremental formalism* [27] is an important aspect of semantic interaction, we must test the models’ ability to learn incrementally over the course of many interactions. To simulate and evaluate the *incremental inference* ability of these models, the selected interactions will be iteratively passed into the SI model. Through these iterative SIs, the underlying models incrementally update to display better results to the analyst.

For example, in the task  $T_{vis}$  (sec. 3.2.2), we simulate that the analyst wants to organize documents by separating the two conferences (InfoVis and VAST) in the 2D projection as ground truth: InfoVis documents placed at top left of the projection, and VAST documents placed at bottom right. Then in each loop, five documents from InfoVis collection, and five documents from VAST collection

Model	$T_{rec}$	$T_{religion}$	$T_{sys}$	$T_{vis}$
$SI_{embedding}$	<b>(0.921, 0.812)</b>	<b>(0.829, 0.773)</b>	<b>(0.895, 0.774)</b>	<b>(0.958, 0.809)</b>
$SI_{keyword}$	(0.497, 0.570)	(0.576, 0.581)	(0.511, 0.584)	<b>(0.961, 0.793)</b>

Table 1: Accuracies of  $SI_{embedding}$  and  $SI_{keyword}$  on each of the four tasks ( $T_{rec}$ ,  $T_{religion}$ ,  $T_{sys}$ , and  $T_{vis}$ ). Two kinds of accuracies are measured: the average accuracy of the trained model after the last interaction and the average accuracy of the models over all interactions.

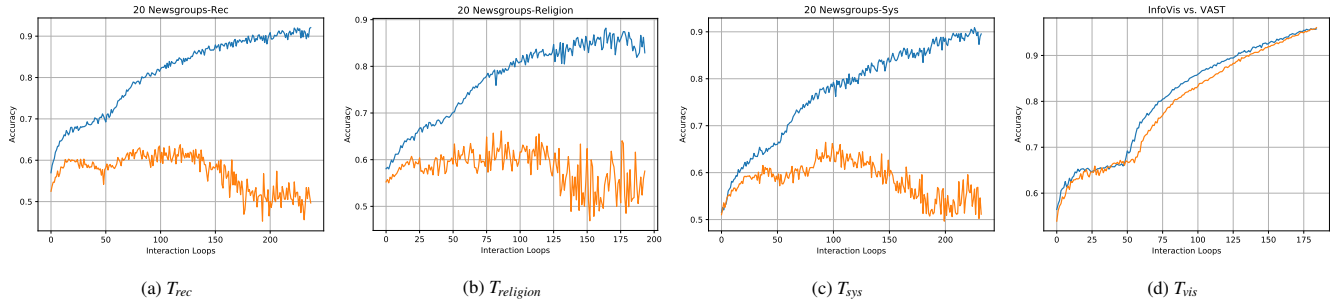


Figure 3: The accuracies of both  $SI_{embedding}$  (blue) and  $SI_{keyword}$  (orange) over each interaction across the four tasks ( $T_{rec}$ ,  $T_{religion}$ ,  $T_{sys}$ , and  $T_{vis}$ ).

will be picked and simulated as semantic interactions: moving five documents from InfoVis to (0,0) position and moving another five documents from VAST to (1,1) position. In each iteration, the SI model ( $SI_{embedding}$  or  $SI_{keyword}$ ) is executed to infer updated model parameters from the simulated semantic interactions. Then the inference ability of each model can be calculated by the quality of the underlying model in classifying InfoVis documents from VAST based on the ground truth.

### 3.2.2 Datasets and Tasks

To perform this quantitative analysis, we use two datasets: 20News dataset [16] and Vispubdata dataset [14]. The 20 Newsgroup dataset is a collection of newsgroups posts on 20 topics. Based on this dataset, we create three tasks to simulate users’ intents. Each of the 3 tasks are to separate the documents into 2 groups based on 2 pre-determined topics:  $T_{rec}$ : 594 documents from “rec.autos” and 600 documents “rec.motorcycles”;  $T_{sys}$ : 578 documents from “comp.sys.mac.hardware” and 592 documents from “comp.sys.ibm.pc.hardware”;  $T_{religion}$ : 379 documents from “talk.religion.misc” and 599 documents from “soc.religion.christian”.

The Vispubdata dataset contains information on IEEE Visualization (IEEE VIS) publications. We select the abstracts of the academic papers published in two conferences from the dataset: InfoVis (IEEE Information Visualization), and VAST (IEEE Visual Analytics Science and Technology). The documents from different conferences are used to represent different concepts (users’ intents). We create one task based on these abstracts to separate InfoVis papers from VAST papers, defined as  $T_{vis}$ , including 397 papers from InfoVis and 531 papers from VAST.

### 3.2.3 Results

In this experiment, we have run the  $SI_{embedding}$  and  $SI_{keyword}$  model multiple times on the four tasks with ground truth. After each interaction (involving five document movements), we calculate the model accuracy using the k-nearest-neighbour (kNN) classifier [8] (as done in Dis-Function [5]): using the cross-validation over the data and set k to 3. The average accuracy of the two models over all iterations for four tasks is shown in Figure 3.

Table 1 shows the average accuracy of the last interaction (performed multiple times) and average accuracy of  $SI_{embedding}$  and  $SI_{keyword}$  across the four tasks.  $SI_{embedding}$  achieved the highest scores in the first three tasks, indicating that  $SI_{embedding}$  provides

more accurate representations of the documents after the interactions, indicating its higher inference ability over  $SI_{keyword}$ . In  $T_{vis}$ ,  $SI_{embedding}$  and  $SI_{keyword}$  have similar accuracies, meaning their inference abilities were roughly equal in this task.

As shown in the Figure 3, we further analyzed the incremental updates from the two models by evaluating each model’s accuracy after every iteration, thereby simulating the incremental analysis process. In the first three tasks, it is shown that using  $SI_{embedding}$  can get better performance than  $SI_{keyword}$  over all the iteration loops. The accuracy of  $SI_{embedding}$  generally increases with each interaction, showing incremental changes to the model’s representation of the user’s intent over time. In contrast to the first three tasks, however, in  $T_{vis}$ , both  $SI_{embedding}$  and  $SI_{keyword}$  have similar accuracy over the interactions. Finally, the desired accuracy should be as close to 1 as possible.  $T_{vis}$  is a relatively “easy” task, as reflected by the final accuracies of both models close to 1.0. These results indicate that  $SI_{keyword}$  performs well in relatively easy tasks.

### 3.3 Discussion

Our experimental results substantiate our hypothesis that  $SI_{embedding}$  is more effective than  $SI_{keyword}$  for modeling user intent, and better supports incremental formalism based on users iterative interactions. The human-centered qualitatively evaluation method shows  $SI_{embedding}$  is more effective than  $SI_{keyword}$  in real world analysis tasks. This provides a direct evaluation of SI models by the observed utility and effectiveness for end-users. The quantitative analysis offers a complementary approach, and provides replicable and scalable evaluations for SI models. It provides more stable and detailed feedback about the SI model performance in tasks of different difficulty levels. Thus, we confirmed that word embedding can better support SI by better capturing the users’ high-level interactive intents.

## 4 CONCLUSION

In this work, we presented  $SI_{embedding}$  as an alternative to the traditional bag-of-words model ( $SI_{keyword}$ ) often used in visual text analytics systems. To make a complete and convincing comparison between  $SI_{embedding}$  and  $SI_{keyword}$ , we performed a quantitative evaluation by simulating analysts’ interactions and calculating the accuracy of the underlying trained ML models, as a complement to the traditional user-centered qualitative evaluation. Results indicate that deep learning distributed representations, such as word embedding, can be exploited to improve interactive visual analytics methods such as semantic interaction.

## REFERENCES

- [1] S. Amershi, J. Fogarty, and D. Weld. Regroup: Interactive machine learning for on-demand group creation in social networks. In *Proceeding CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, May 2012.
- [2] B. Athiwaratkun and K. Kang. Feature representation in convolutional neural networks. *arXiv preprint arXiv:1507.02313*, 2015.
- [3] N. Boukhelifa, A. Bezerianos, and E. Lutton. Evaluation of interactive machine learning systems. In *Human and Machine Learning*, pp. 341–360. Springer, 2018.
- [4] L. Bradel, C. North, L. House, and S. Leman. Multi-model semantic interaction for text analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 163–172, Oct 2014. doi: 10.1109/VAST.2014.7042492
- [5] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 83–92, Oct 2012. doi: 10.1109/VAST.2012.6400486
- [6] R. Collobert and J. Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pp. 160–167. ACM, 2008.
- [7] K. A. Cook and J. J. Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.
- [8] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.*, 13(1):21–27, Sept. 2006. doi: 10.1109/TIT.1967.1053964
- [9] M. Dowling, N. Wycoff, B. Mayer, J. Wenskovitch, S. Leman, L. House, N. Polys, C. North, and P. Hauck. Interactive visual analytics for sensemaking with big text. *Big Data Research*, 16:49 – 58, 2019. doi: 10.1016/j.bdr.2019.04.003
- [10] A. Endert, R. Chang, C. North, and M. Zhou. Semantic Interaction: Coupling Cognition and Computation through Usable Interactive Analytics. *IEEE Computer Graphics and Applications*, 35(4):94–99, jul 2015. doi: 10.1109/MCG.2015.91
- [11] A. Endert, P. Fiaux, and C. North. Semantic interaction for sensemaking: inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, 2012.
- [12] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pp. 473–482. ACM, New York, NY, USA, 2012. doi: 10.1145/2207676.2207741
- [13] F. Hughes and D. Schum. Discovery-proof-choice, the art and science of the process of intelligence analysis-preparing for the future of intelligence analysis. *Washington, DC: Joint Military Intelligence College*, 2003.
- [14] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. Stolper, M. Sedlmair, J. Chen, T. Möller, and J. Stasko. vispubdata.org: A metadatabase collection about IEEE visualization (VIS) publications. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2199–2206, Sept. 2017. doi: 10.1109/TVCG.2016.2615308
- [15] M. J. Kusner, Y. Sun, N. I. Kolkin, and K. Q. Weinberger. From word embeddings to document distances. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, pp. 957–966. JMLR.org, 2015.
- [16] K. Lang. Newsweeder: Learning to filter netnews. In *Proceedings of the Twelfth International Conference on Machine Learning*, pp. 331–339, 1995.
- [17] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, ICML'14*, pp. II–1188–II–1196. JMLR.org, 2014.
- [18] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [19] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [20] 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*, pp. 3111–3119, 2013.
- [21] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.
- [22] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep contextualized word representations. In *Proc. of NAACL*, 2018.
- [23] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. pp. 2–4, 2005.
- [24] J. Z. Self, M. Dowling, J. Wenskovitch, I. Crandell, M. Wang, L. House, and C. North. Observation-level and parametric interaction for high-dimensional data analysis. *ACM Transactions on Interactive Intelligent Systems (TiS)*, 8(2):15, 2018.
- [25] J. Z. Self, R. K. Vinayagam, J. Fry, and C. North. Bridging the gap between user intention and model parameters for human-in-the-loop data analytics. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, p. 3. ACM, 2016.
- [26] F. Shaheen, B. Verma, and M. Asafuddoula. Impact of automatic feature extraction in deep learning architecture. In *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, Nov 2016. doi: 10.1109/DICTA.2016.7797053
- [27] F. M. Shipman and C. C. Marshall. Formality considered harmful: Experiences, emerging themes, and directions on the use of formal representations in interactive systems. *Computer Supported Cooperative Work (CSCW)*, 8(4):333–352, Dec 1999. doi: 10.1023/A:1008716330212
- [28] R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng. Parsing with compositional vector grammars. In *In Proceedings of the ACL conference*, 2013.
- [29] J. Yosinski, J. Clune, Y. B. A. i. neural, and 2014. How transferable are features in deep neural networks? *papers.nips.cc*.
- [30] T. Young, D. Hazarika, S. Poria, and E. Cambria. Recent trends in deep learning based natural language processing [review article]. *IEEE Computational Intelligence Magazine*, 13(3):55–75, Aug 2018. doi: 10.1109/MCI.2018.2840738
- [31] Y. Zhang, R. Jin, and Z.-H. Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52, 2010.