Using Natural Language Processing Techniques to Tackle the Construct Identity Problem in Information Systems Research

Siegfried M. Ludwig
Leuphana University Lueneburg
siegfried.m.ludwig@gmail.com

Burkhardt Funk
Leuphana University Lueneburg
funk@uni.leuphana.de

Benjamin Mueller
University of Lausanne
benjamin.mueller@unil.ch

Abstract

The growing number of constructs in behavioral research presents a problem to theory integration, since constructs cannot clearly be discriminated from each other. Recently there have been efforts to employ natural language processing techniques to tackle the construct identity problem. This paper compares the performance of the novel word-embedding model GloVe and different document projection methods with a latent semantic analysis (LSA) used in recent literature. The results show that making use of an advantage in document projection that LSA has over GloVe, performance can be improved. Even against this advantage of LSA, GloVe reaches comparable performance, and adjusted word embedding models can make up for this advantage. The proposed approach therefore presents a promising pathway for theory integration in behavioral research.

1. Introduction

The metaphor of standing on the shoulders of giants has long been recognized as a powerful image to argue that good research needs to be grounded in extant literature to ensure that new contributions take into account what is already known and sufficiently advance the state-of-the-art [1]. To be done well, these essential tasks require scholars aspiring to contribute to an ongoing discourse on their phenomenon to be well erudite of the pertinent literature. Mirroring this requirement, the Information Systems (IS) discipline has invested substantial efforts into building up methodological guidance on how to do literature reviews (e.g., [2, 3, 4]).

While this guidance is invaluable in crafting literature reviews, the discipline's renewed engagement with what it already knows has also led to an intensified recognition of the challenges inherent in working with extant literature. Chief among them, for the purpose of our work, is the construct identity problem (e.g., [5, 6]). It suggests that identifying specific phenomena a study refers to is complicated because researchers are essentially free to name constructs as they see fit. This results in a largely unstandardized and often conflicting set of names used to refer to the various phenomena of interest for IS research.

This makes identifying whether a study one finds when searching for a specific construct's seemingly accepted name actually provides pertinent insight rather complicated and effortful. For instance, imagine researching users post-adoption resistance behaviors. Looking at the extant literature will return multiple papers, with Lapointe and Rivard [7] and Bala and Venkatesh [8] likely among them. Both papers use a construct called perceived threat, but can we be sure that both mean the same thing that is, an identical conceptual entity with the very same definition, scope conditions, and semantic relationships (as the building blocks of construct clarity suggested by Suddaby [9])? As highlighted by Boell and Cecez-Kecmanovic [3], especially allegedly 'systematic reviews' using string-based database research often overlook synonyms used to refer to the phenomenon of interest. At the same time, there is a risk that such research erroneously includes research covering a different phenomenon, just because the authors there employed the same name for their construct two problems know as jangle fallacy [10] and jingle fallacy [11] respectively.

If such construct identity problems occur, an adequate representation of the literature becomes challenging. This is especially true for studies that seek to identify relevant intellectual overlap in a discipline's theoretical findings or that aim at consolidating what is known about a specific phenomenon (e.g., [12]).

As one possible remedy to this problem, beyond meticulous reading of the literature and the effortful accomplishment of exhaustive erudition in a study's subject-matter domain, the literature has recently begun to explore machine-learning-based approaches to support theorists (esp. [6]). While this research provides an important impulse and first artifacts to support scholars struggling with construct identity

HĭCSS

problems, respective studies have thus far mainly relied on latent semantic analysis (LSA). We propose that the field of machine learning has progressed since and that especially novel approaches like GloVe offer promising opportunities to advance research into automated support to analyzing a construct's identity.

2. Related work

Developing our approach, we first introduce the theoretical problem - construct identity - and its implication for research work. We then look at existing approaches before introducing recent advances in machine learning that form the basis for our approach.

2.1. Construct identity challenges

In most of behavioral research in IS, constructs play a central role [1]. In general terms, researchers use constructs as conceptual abstractions of classes of observations [13]. Constructs thus serve as heuristic devices for researchers to make sense of the empirical world and help to combine what is learned from individually observed instances of a phenomenon [14]. Consequently, constructs can be thought of as central vessels for documenting and communicating knowledge about a phenomenon.

Given this centrality of constructs in research, especially from a cumulative point-of-view, the ability to correctly identify constructs is paramount. Consider, for instance, a construct such as 'deep structure'. On the one hand, studies that relate to Wand and Weber's [15] representation theory use the concept to describe parts of an information system users do not generally interface with. On the other hand, Gersick [16] uses the same term to refer to deeply ingrained and institutionalized characteristics of organizations. While scrutiny of the terms will reveal that these two instances are conceptually unrelated - for instance when applying Suddaby's [9] concept of construct clarity - a confusing duplication of the principal name remains.

In the literature, challenges like this have been summarized under the term 'construct identity problems' and receive growing recognition (e.g., [5, 6]). In particular, two variants of the construct identity problem are discussed:

First, the *jingle fallacy* describes the occurrence of two constructs with identical names but referencing different real-world phenomena [11]. This is illustrated by the deep structure example we use above.

Second, the *jangle fallacy* occurs when different construct names are being used to refer to the same real-world phenomenon [10]. What results is the proliferation of a large number of seemingly new constructs simply based on a naming problem. It has been shown that the probability of different people, even experts, choosing the same word to describe a concept is less than 20% [17]. This means that if people do not know the name of a construct they are searching for in advance, they are very unlikely to find it. As pointed out by Larsen and Bong [6], this results in researchers missing the majority of papers that employ constructs relevant to them.

When present, we suggest that the jingle and jangle fallacy greatly increase the likelihood of type I and type II errors in working with the literature we discussed earlier. Similarly, especially the jangle fallacy increases the likelihood of researchers thinking that they did coin a new construct, even though they are essentially reinventing the wheel. This adds to the fragmentation of knowledge IS research is increasingly criticized for (e.g., [5, 1, 12]).

In effect, the standing on the shoulder of giants we invoked earlier becomes increasingly difficult. To avoid standing on the shoulders of an illusory giant made of incompatible pieces or to suspect a greenfield where there is none, researchers need to carefully identify what is known about a phenomenon.

2.2. Current approaches

To date, the discovery processes needed to learn what is known about a phenomenon have been mainly looked at from a purely methodological point of view - recent discussions on literature reviews being a case in point (e.g., [2, 3, 4]). In these, however, problems related to construct identity do not seem to receive either explicit nor widespread attention. Nonetheless, we propose that there are three main approaches to tackling construct identity problems in IS research that need to be acknowledged. To do so, we briefly introduce and discuss the respective idealistic foci below.

The first approach is focused on what Rivard [18] calls *erudition*, which refers to "[...] the breadth and depth of ones knowledge of a particular topic" (p. vii). In idealistic terms, this refers to scholars well-versed in their domain and aware of all the pertinent literature through years of reading and participating in the scholarly discourse. Arguably, erudite scholars have such a profound knowledge of their domain that they will be implicitly aware of synonyms used for a relevant construct and homonyms they can competently ignore. While perhaps the most noble form of being able to identify and avoid or resolve construct identity problems, the major drawback of this approach is that it is enormously effortful and often requires entire careers to achieve. It is further complicated by the increasing

volume of scientific publications [19].

These drawbacks have given rise to the second approach, focused on *structured or systematic ways* to review the literature. If the well-versed senior scholar represented the first approach, the second approach is likely represented by hesitant PhD students faced with the daunting task of getting up to speed in their relevant and reasonably related discourses. These studies employ structured, protocol-based reviews of the literature (e.g., [4, 20, 2]). If done carefully, this approach produces a comprehensive and intimate representation of the literature, exploring hidden knowledge about a phenomenon that adds on top of the current state-of-the-art (e.g., [21, 22]). However, recent critique illustrates that such mainly mechanistic approaches remain quite susceptible to construct identity problems because they are often based on a false sense of systematicity, especially when simply string-based search approaches are used [3]. While the literature suggests remedies - such as corroboration by bibliometric approaches like the forward and backward search proposed by Webster and Watson [2] - this adds substantial effort to the otherwise rather efficient approach. While tool support for this activity is discussed in the literature [23], this mostly pertains to record keeping and support of content analysis similar to tool support employed in the analysis of primary qualitative empirical data.

In response to the respective strengths and weaknesses of the two approaches discussed thus far, research has more recently begun to discuss *tool-based* approaches. Similar to computer-aided work with unstructured and unstandardized documents in the context of requirements management [24, 25], tool-based approaches aim to reduce the efforts and cognitive load of the construct identification task by training a machine to identify and classify constructs. In particular, Larsen and Bong [6] suggest a machine learning approach based on LSA. Such hybrid approaches likely outperform purely structured approaches because they do not only rely on simply string-based keywords, but on an actual semantic analysis. This helps to overcome problems that arise from (either explicit or implicit) disciplinary naming conventions. At the same time, once set up, such a tool-based approach is much more time efficient than building up erudition alone (the preparatory efforts to set up the algorithm and database notwithstanding).

Larsen and Bong [6] demonstrate that the LSA based approach shows significant predictive power in the construct identification task, while, at the same time, not being able to fully match experts' abilities in this task (AUC=0.791). Having said that, we acknowledge that the field of machine learning has made methodological progress since the work by Larsen and Bong. Thus, we suggest that other, rivalry methods available for automated natural language processing (NLP) might help tool-supported approaches to tackling construct identity problems in IS research.

2.3. Methodological progress

In order to implement tool-based approaches two constituents are needed: First, an extensive dataset of articles containing the constructs of interest has to be set up. From these articles various elements need to be extracted either manually or supported by a tool. These elements include construct names and might encompass other data such as definitions of constructs, indicators, model parameters, meta data of the articles and free text. Second, tool-based approaches need to have a way to determine the similarity or identity of constructs. Only then these tools can support theoretical research in IS, for instance by finding articles that discuss a specific construct of interest.

For the first constituent, Larsen and Bong [6] generated a dataset by manually categorizing 1004 constructs from 193 articles from the journals MIS Quarterly and Information Systems Research into 347 clusters of correspondent constructs. Constructs are defined as correspondent if they refer to the same real-world phenomenon. The resulting dataset is publicly available and also used in our analysis [26].

For the second constituent, Larsen and Bong [6] used LSA to automatically determine the similarity of different constructs based on the semantics of the construct indicators (for a discussion of their CID1 ensemble, refer to section 6). From a theory standpoint LSA makes use of the distributional structure of language as it can be found in statistics of word co-occurrences. This kind of model is based on the observation that the environment in which a given word tends to occur yields information about the meaning of the word [27]. By analysing word co-occurrences and placing those words into a high-dimensional space accordingly, machine learning models can learn relationships between those words.

These relationships can be used to determine the similarity of different constructs based on the semantic relatedness of their constituent indicators. An indicator, usually a question asked to study subjects, consists of multiple words. In order to compare items with each other, they have to be projected into the previously learned semantic space. Different document projection methods are available for this.

From a theoretical standpoint, semantic similarity

measures between different constructs open up interesting possibilities. Larsen et al. [28] draw parallels to the multitrait-multimethod matrix, specifically convergent and divergent validity. The former states that the degree of correlation between measures of the same construct should be high, while the latter states that measures of other constructs should not be correlated. They argue that the LSA similarity measure, computed with LSA, can help to assess these two factors of construct validity on a semantic basis.

Developed in the 1990s [29], LSA, however, has been extended and complemented by novel machine learning approaches. In particular, methods for word representation or so-called word embeddings [30][31] show promising results going beyond what can be achieved with LSA. One feature that makes them stand out against LSA is the improved structure of the vector space with linear relationships between vectors (e.g. king - man + woman = queen).

3. Method

Against this backdrop, we pursued a research approach inspired by design research (esp. [32]). Our initial problem awareness roots equally in our own experiences as IS researchers as well as the increasing attention dedicated to problems connected to construct identity in the IS literature [5, 6, 12, 1]. As demonstrated by Larsen and Bong [6], tool-based approaches based on machine learning seemed a promising approach to tackle construct identity problems. As indicated earlier, we felt that this course of action is further supported by similar advances in computer-aided work with unstructured documents elsewhere [24, 25].

In the development of our specific artifact, we are guided by the ideal that a design problem is not solved simply by the first functioning artifact that proves viable. We rather feel inspired by Weick's [33] suggestion of variation, selection, and retention as guiding principles for theoretical progress in a discipline. We thus set out to explore alternative solutions that seek to extend and enhance the solution proposed by Larsen and Bong [6]. As discussed earlier, the methodological progress achieved in the domain of machine learning further contributed to our intention to advance Larsen and Bong's [6] line of thinking.

For the demonstration of our artifact, the results achieved with LSA [6] on the construct identity task are reproduced and serve as a baseline to evaluate our proposed GloVe-based solution. Different approaches to document projection are explored and compared.

4. A GloVe-based solution

In section 2.3 we discussed the two constituents of tool-based approaches. While we use the dataset of Larsen and Bong [6] for our analysis to enable a comparison of results, here we focus on the second constituent, that is implementing similarity measures for constructs. To determine word and subsequently item and construct similarity we propose to employ novel word embedding techniques, specifically GloVe introduced by Pennington et al. in 2014 [31]. We expect GloVe to outperform LSA when implementing similarity measures, which is supported by the work of Baroni et al. [34], who conducted a systematic comparison of traditional count-based models, including LSA, and prediction-based embedding models, including word2vec. Looking at more than ten different tasks they found that novel word embeddings outperform traditional count-based models on almost all tasks. Of specific interest for this paper is the performance on tests for semantic relatedness and synonym detection, on which prediction-based models outperformed count-based models by about 10 percentage points. Their tests did not include GloVe, however, since GloVe features the same linear relationships between concepts in its semantic space as word2vec and since Pennington et al. claim the performance to be comparable to that of word2vec [31], GloVe should be able to outperform LSA on the construct identity detection task. Also Naili et al. found that GloVe performs equally well with word2vec and both models are superior in performance to LSA [35].

In this section we briefly review the preprocessing steps in NLP needed to enable machines to reason on natural language. We then contrast the differences of LSA and GloVe. Ultimately, we explain how GloVe is used to implement a construct similarity measure in our solution.

4.1. Preprocessing in NLP

Since natural language is highly unstructured, it has to be preprocessed to make it machine-readable. A number of steps are typically applied [36]: words are converted to lower-case and certain non-textual characters like commata and brackets are removed. Stop-words like the in and it, which frequently appear and carry almost no context-specific meaning, are removed as well. Then stemming can be performed, which removes the suffixes and prefixes from each word, and, depending on the stemming algorithm, can try to reconstruct a valid word from the stem. The rule-based stemming method Porter proposed by Porter [37] and

later improved as Porter2 [38], is still very popular today and will be used in this analysis.

As a next processing step, a document-term co-occurrence matrix is created, which counts how often each word occurs in each document. Since some words appear more often in a document than others, but have little meaning to the context, a weighting method is usually used. A popular weighting factor is the term frequency-inverse document frequency (TF-IDF). With TF-IDF, document word counts are penalized by how often the word appears in the training corpus, which can give less frequent and therefore more relevant words more impact. Larsen and Bong [6] used log-entropy weighting instead of TF-IDF. There is some literature that suggests that log-entropy weighting can outperform TF-IDF [39, 40], but TF-IDF is more popular. In this analysis we compare both approaches. After weighting the co-occurrence matrix, it is normalized using the l^2-norm.

4.2. Latent Semantic Analysis (LSA)

LSA creates a semantic space of specifiable dimensionality, in which associated words and documents are placed close to each other. The cosine similarity between the word or document vectors can then be taken as their respective semantic similarity. To create this semantic space and measure word and document similarity, the document-term co-occurrence matrix is used as a starting point. LSA finds a low-rank approximation of this matrix by first decomposing it into singular values and vectors. This representation is then reduced to the specified dimensionality by removing the weaker components [41]. A good value for the number of dimensions has been established to be around 300 [41], which is the number used by Larsen and Bong [6] and consequently also in our analysis. The number of dimensions should neither be too low nor too high, but a specific value is hard to establish theoretically and might vary by the context of application.

4.3. GloVe

The original paper on GloVe [31] appeared shortly after Mikolov et al. [42] published their impactful paper proposing word2vec, which Pennington et al. [31] make reference to, and its follow-up improvements [30]. Word2vec uses a shallow neural network related to the neural language model proposed by Bengio et al. [43]. There are two approaches to word2vec, distinguished by their training objective. While Continuous Bag Of Words (CBOW) tries to learn word vectors that are good in predicting a central word given the nearby

words, the skip-gram version tries to learn word vectors that are good at predicting the surrounding words from a central word. Each of these approaches has preferred application contexts where they outperformed each other, with CBOW being the strongest for syntactic similarity and skip-gram being the strongest for semantic similarity [42].

GloVe was then proposed by Pennington et al. [31] as a way to combine preferable characteristics of global matrix factorization methods like LSA, which make efficient use of global statistical information, and local context window methods like word2vec, which have a better vector space structure, indicated by the better performance on analogy tasks. Word2vec provides linear relationships between concepts in its vector space, allowing for arithmetic like $king - man + woman = queen$ [42]. This feature is implemented in GloVe and gives it a superior vector space structure over LSA, while keeping the use of global corpus statistics as in LSA. Word2vec performs an implicit matrix factorization [44], which was made explicit with GloVe in addition to other adjustments [45].

To perform an analysis with GloVe, first a sparse term co-occurrence dictionary is created from the text corpus. GloVe is then trained on that dictionary and creates word vectors of the specified dimensionality. In our analysis we use 300 dimensions, the same number as is used for LSA. Further hyperparameters available for tuning include the learning rate, alpha and x_max, of which the latter two parameterize the function for weighting less frequent co-occurrences. It is to be expected that the ideal hyperparameters for this context can differ from those reported by Pennington et al. [31], because here we are training on only around 5000 short item texts instead of larger text corpora as is usually done. For our analysis we use a learning rate of 0.2, alpha of 0.4 and x_max of 10, training for 50 epochs.

4.4. Document projection

If only word vectors are available, documents have to be projected into the semantic space of those word vectors. A document is here defined as a collection of words, which can be a sentence, a paragraph or a full text. In this particular use case LSA has the advantage that when trained on the construct items directly, singular value decomposition creates both word vectors and also document vectors for those items. Since in this case LSA can be trained on the corpus under scrutiny, document projection is not an issue. The GloVe implementation used for this analysis does not provide document vectors, however. For their analysis Larsen and Bong [6] trained LSA on paragraphs of research

papers and therefore did not have document vectors of the construct items. They used some document projection method, but did not provide information about the details. In our analysis all LSA and GloVe setups are trained directly on the construct items of the gold standard dataset. This allows us to use the LSA-trained document vectors and it results in a smaller training set, allowing for shorter training times.

A popular method for document projection is taking the weighted or unweighted centroid (average) of the word vectors making up the document. Afterwards the document vectors are in both cases normalized again. The document similarity can then simply be defined as the cosine similarity between the document vectors. The literature suggests that the weighted centroid should outperform the unweighted one [46].

We compute the unweighted centroid as the simple average of the vectors of unique words in the item. The weighted centroid is computed by taking the sum of the word vectors making up a document, where each word vector is multiplied by its l^2 normalized TF-IDF value in the document-term matrix, and then dividing by number of unique words in the document.

4.5. Construct similarity

To go from item vectors to construct similarity, Larsen and Bong [6] computed cosine similarities of the item vectors between two constructs. They then took the average of the similarity of the most similar item pair scores to be the similarity of the two constructs. However, it is not clear to us how many item pair similarities they included in the average. We chose to work with the two highest similarities. They write that [t]his function tends to compensate for item pairs that are constituted of both highly similar and very dissimilar constituent parts. ([6], p. A16).

We implemented the solution in Python using standard libraries such as scikit-learn, numpy and stemming. For GloVe, a publicly available implementation is used [47]. The code is available upon request to the authors.

5. Results

A popular performance measure in binary classification tasks, also employed by Larsen and Bong [6], is the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. The ROC curve plots the true positive rate against the false positive rate. It is the main performance measure employed in this analysis.

Comparing the results of different co-occurrence weighting and document projection methods showed that TF-IDF generally slightly outperformed log-entropy, albeit not by much (see table 1). The best results for LSA were achieved with TF-IDF weighting and using the LSA-trained document vectors (construct items), circumventing the document projection problem. Larsen and Bong achieved a performance baseline of 0.791 for LSA, using some document projection method [6]. Since the document projection method they used is unknown, the results could not be perfectly compared to ours. What is clear from our results however is that the model performance can be improved by just training on the items directly and using the provided document vectors. In summary, directly LSA-trained document vectors outperform unweighted and weighted centroids.

Table 1. Performance of document projection methods on LSA (ROC AUC).

ROC AUC	document vectors	unweighted centroid	weighted centroid
log-entropy	0.815	0.816	-/-
TF-IDF	**0.823**	0.815	0.792

Table 2 presents the correlations of the item similarities computed with the three different document projection methods. Interestingly, although the different methods achieve a performance very close to each other, they are not fully correlated. This suggests that different document projection techniques might be able to contribute unique information and could work together well in an ensemble, which is a question for further research.

Table 2. Correlations between document projection methods on LSA.

	doc-vectors	unw-cent	w-cent
doc-vectors	1.0		
unw-cent	0.902	1.0	
w-cent	0.879	0.906	1.0

The comparison of LSA and GloVe resulted in almost equal performance, with GloVe having a slight edge (see table 3, figure 1). The best configuration for LSA was with TF-IDF weighting and using LSA-trained document vectors (0.8231). GloVe achieved the best results when trained on a term co-occurrence dictionary based on a TF-IDF-weighted document-term matrix and using unweighted vector centroids for document projection (0.8274). That LSA contrary to expectation achieved almost the same result as GloVe can be explained with the document projection problem. Since directly trained document vectors perform better than vector centroids, using GloVe term vectors comes with a

disadvantage. When both models use unweighted vector centroids, GloVe (0.8274) outperforms LSA (0.8157) by a somewhat larger margin.

Table 3. Performance of LSA and GloVe (ROC AUC).

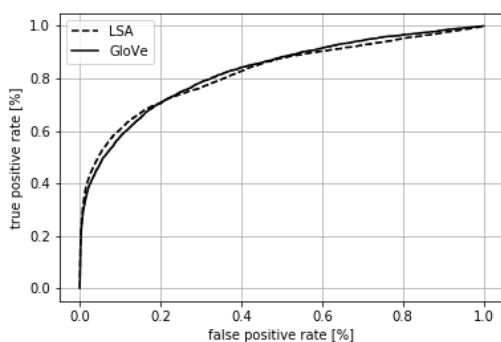| model	ROC AUC
LSA reported baseline [6]	0.791
LSA TF-IDF w-cent	0.792
LSA TF-IDF lsa-docs	0.823
GloVe TF-IDF unw-cent	**0.827**

Figure 1. Performance of LSA with TF-IDF weighting and directly created document vectors and GloVe with TF-IDF weighting and unweighted centroid document projection (ROC curve).

The correlation between construct similaritiy scores determined by LSA and GloVe is 0.774. Even more than with the correlations between the document projection methods, the outputs of the models correlate less with each other than expected. Again, this could mean that they would contribute unique information to an ensemble, which could be looked into in future research.

6. Discussion

It is worth pointing out that both LSA and GloVe had very good performance even when only trained on the item texts directly as opposed to being trained on texts of research papers as done by Larsen and Bong. This might mean that although our training corpus is much smaller, it is more relevant to the problem at hand and therefore leads to very comparable performance. The word embedding model might therefore learn the semantic relationships between items more directly and clearly, instead of relying on semantic relationships observed in general language patterns.

Even with the advantage of directly providing document vectors that LSA has over GloVe, the proposed method was able to perform at least equally

well. The analysis conducted with LSA suggests that, although the items are very short documents, directly trained document vectors tend to perform better on this task than document vectors aggregated from word vectors. The GloVe implementation does not provide document vectors for the construct items and some of the performance gets lost during document projection. The fact that GloVe nevertheless provides at least equally good results shows the potential of novel word embedding models. Word2vec has extensions that allow for the direct training of document vectors. Thereby the advantage of LSA over embedding models can be equalized and superior performance of word2vec is to be expected. We therefore propose to use word2vec in future research. Word2vec extensions for document vectors include doc2vec [48], doc2vecC [49] and sent2vec [50].

In addition to the advantage word2vec provides over GloVe with regards to document vectors, contrary to Pennington et al. [31], GloVe has in some papers been shown to be inferior to word2vec when both models are trained with optimized hyperparameters [45]. An analysis with word2vec could therefore improve performance on the construct identity detection task, even beyond the document projection advantages.

Larsen and Bong [6] achieved their best results when adding a knowledge-based model, which makes use of length and depth distances between words in an existing knowledge base [51], in an ensemble with LSA and named it Construct Identity Detector (CID1). Using this ensemble improved the predictive performance to some degree (AUC = 0.814). Given that the correlation between LSA and GloVe of 0.774 is rather low we expect an ensemble learner that include LSA and GloVe to further improve the performance.

In the above analysis, the construct similarities are obtained by aggregating semantic similarities between the constituent items. Different options to get vectors of constructs projected into the semantic space include directly taking a centroid of the word vectors making up the items of a construct, instead of using the current two-step approach of first creating document vectors for the items and then aggregating the item similarities to the construct similarity. This essentially compares to creating a document vector for a paragraph based on word vectors of all the words in the paragraph, instead of based on vectors of the sentences in the paragraph. Construct vectors could also be created by taking a centroid of the item vectors, instead of the aggregation method taking the average similarity of the most similar items proposed by Larsen and Bong [6]. The construct similarity could then in both cases be calculated by taking a simple cosine similarity between the construct

vectors. These are some possibilities for future research.

While these potentials for future research illustrate important aspects that additional research and development efforts should be directed to, our work nonetheless shows the potential that tool-based approaches to tackling the construct identify problem in IS research have. Especially when done on the basis of comprehensive, document-based analyses, machine-learning-based approaches have the potential to overcome the limitations of traditional approaches. More specifically, they are likely to outperform simple structured or systematic efforts to identify relevant extant work in terms of accuracy and comprehensiveness. While we did not yet evaluate our artifact in comparison with traditional approaches, the results obtained in other domains provide encouraging evidence to suspect as much [24]. Similarly, we think it is reasonable to assume that especially novice scholars that want to get a reliable and comprehensive overview of a specific domain will find tool-based approaches more efficient than pure erudition-based approaches.

Reflecting on our findings on a more abstract level also leads us to observe that a collection of key words does not contain the full semantic content of a concept. This is not only due to the jingle and jangle fallacies, but also apparent in the superior performance of construct items that are directly projected into vector space over item vectors that are some combination of word vectors. String-based literature search might suffer from a similar problem: the underlying semantics of the concept a researcher is looking for are not fully captured by using a collection of key words. A tool like the one proposed in this paper can help researchers by enabling them to make use of semantic relationships between research constructs, which can be more useful and accurate than string-based approaches.

Long term, we think that employing machine-learning based tools to help with construct identification and analysis also bears potentials for advances in how we do theoretical work altogether. Especially in settings where we find that a machine approaches the problem differently than its human counterpart, the reflective integration of the results promises to be quite insightful and instructive for the theorist.

7. Conclusion

We proposed to use novel word embedding models for improving the automated semantics-based approach to theory integration developed by Larsen and Bong [6] and demonstrate promising results. The analysis of different document projection methods yielded useful insights into an advantage of LSA that led us to propose the investigation of the word-embedding model word2vec in future research.

Tool-based approaches to theory integration do not reach human performance yet when it comes to determining the identity of two given constructs. While this accuracy can be increased with future research, those tools already have the promising advantage of allowing to relate constructs on a semantic basis. This can start to tackle the problem of systematic reviews that are mostly limited to search by keywords and thereby can miss considerable amounts of research, as has been criticized by Boell and Cecez-Kecmanovic [3].

The promises of tool-based approaches notwithstanding, we would like to highlight a few limitations of the current state of research. First, the usefulness of the proposed tool-based approach directly depends on the availability and scale of structured databases containing the constructs proposed in behavioral research. To this end, theorists currently can draw on a number of sources. For example, the Inter-Nomological Netowrk [52] provides researchers with a repository of constructs and associated variables. Especially in the context of operationalization, such a repository provides important input for machine-learning based analyses. Beyond this, the current DISKNET project [53] provides both an analytical and a semantic perspective, allowing for an analysis for relations *between* constructs as well. Following Suddaby [9], this is an important extension to fully grasp construct clarity and bears great promises for theory integration. However, prior approaches to build taxonomies or folksonomies in IS research had very little impact (e.g., [54]) such that considerable thinking will have to be invested in how to make the knowledge embedded in the many IS manuscripts published every year accessible for synthesis and scrutiny. Second, as hinted to above, evaluations of our work thus far are limited to how our artifact performs against an established gold standard. Further tests will have to reveal how well tool-based approaches in general perform against the more traditional approaches to tackling the construct identity problem. Third, like with most artifact-based solutions, whether or not tool-based discovery and analysis of constructs will have an impact in IS research will depend on individual and disciplinary adoption. A discussion of the inherent dynamics of adoption and post-adoption of our tool is, however, beyond our scope here.

Nonetheless, we continue to believe that tool-based solutions bear great potential to advance the work of theorists - in IS and beyond. Much like in other domains - such as computer-aided design, computer-aided

software engineering, or computer-aided requirements elicitation - machines bring distinct strengths to the table. However, we would like to highlight that good theoretical work in the future is not something that is entrusted to machines only. We think it far more likely that finding good ways to combine the strengths of an erudite and creative scholar with that of a well-trained tool is the most promising way forward. Only if we learn to race with the machine rather than against them will we become able to reach gigantic heights in our scholarly work.

References

[1] B. Mueller and N. Urbach, "Understanding the why, what, and how of theories in is research.," *CAIS*, vol. 41, p. 17, 2017.

[2] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, pp. xiii–xxiii, 2002.

[3] S. K. Boell and D. Cecez-Kecmanovic, "On being 'systematic' in literature reviews in is," *Journal of Information Technology*, vol. 30, no. 2, pp. 161–173, 2015.

[4] C. Okoli, "A guide to conducting a standalone systematic literature review," *Communications of the Association for Information Systems*, vol. 37, 2015.

[5] A. Rai, "Editor's comments: The first few pages," *MIS Quarterly: Management Information Systems*, vol. 42, pp. iii–vi, 06 2018.

[6] K. R. Larsen and C. H. Bong, "A tool for addressing construct identity in literature reviews and meta-analyses.," *Mis Quarterly*, vol. 40, no. 3, pp. 529–551, 2016.

[7] L. Lapointe and S. Rivard, "A multilevel model of resistance to information technology implementation," *MIS Quarterly*, vol. 29, no. 3, pp. 461–491, 2005.

[8] H. Bala and V. Venkatesh, "Adaptation to information technology: A holistic nomological network from implementation to job outcomes," *Management Science*, vol. 62, no. 1, pp. 159–179, 2016.

[9] R. Suddaby, "Editor's comments: Construct clarity in theories of management and organization," *Academy of Management Review*, vol. 35, no. 3, pp. 346–357, 2010.

[10] T. L. Kelley, *Interpretation of educational measurements.* World Book Co., 1927.

[11] E. L. Thorndike, *An Introduction to the Theory of Mental and Social Measurements.* New York: Science Press, 1904.

[12] D. S. Hovorka, K. R. Larsen, J. Birt, and G. Finnie, "A meta-theoretic approach to theory integration in information systems," in *2013 46th Hawaii International Conference on System Sciences*, pp. 4656–4665, IEEE, 2013.

[13] R. L. Priem and J. E. Butler, "Tautology in the resource-based view and the implications of externally determined resource value: Further comments," *Academy of Management Review*, vol. 26, no. 1, pp. 57–66, 2001.

[14] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*. McGraw-Hill Series in Psychology, New York, NY, USA: McGraw-Hill, 3. ed., 1994.

[15] Y. Wand and R. Weber, "On the deep structure of information systems," *Information Systems Journal*, vol. 5, no. 3, pp. 203–223, 1995.

[16] C. J. G. Gersick, "Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm," *The Academy of Management Review*, vol. 16, no. 1, pp. 10–36, 1991.

[17] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, "The vocabulary problem in human-system communication," *Communications of the ACM*, vol. 30, no. 11, pp. 964–971, 1987.

[18] S. Rivard, "Editor's comments - the ions of theory construction," *MIS Quarterly*, vol. 38, no. 2, pp. iii–xiii, 2014.

[19] L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, pp. 2215–2222, 2015.

[20] G. Schryen, "Writing qualitative is literature reviews–guidelines for synthesis, interpretation and guidance of research," *Communications of the AIS*, vol. 37, no. Art 12, pp. 286–325, 2015.

[21] K. Ebner, N. Urbach, and B. Mueller, "Exploring the path to success: A review of the strategic it benchmarking literature," *Information & Management*, vol. 53, no. 4, pp. 447–466, 2016.

[22] S. S. Li and E. Karahanna, "Online recommendation systems in a b2c e-commerce context: a review and future directions," *Journal of the Association for Information Systems*, vol. 16, no. 2, p. 72, 2015.

[23] W. Bandara, E. Furtmueller, E. Gorbacheva, S. Miskon, and J. Beekhuyzen, "Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support," *Communications of the Association for Information Systems*, vol. 37, pp. 154–204, 2015.

[24] H. Meth, B. Mueller, and A. Maedche, "Designing a requirement mining system," *Journal of the Association for Information Systems*, vol. 16, no. 9, p. 799, 2015.

[25] R. E. Vlas and W. N. Robinson, "Two rule-based natural language strategies for requirements discovery and classification in open source software development projects," *Journal of Management Information Systems*, vol. 28, no. 4, pp. 11–38, 2012.

[26] K. R. Larsen and C. H. Bong, "Gold standard dataset of construct identities." https://www.researchgate.net/publication/317901676_GoldStandard, 2016.

[27] Z. S. Harris, "Distributional structure," *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.

[28] K. R. Larsen, J. Lee, J. Li, and C. H. Bong, "A transdisciplinary approach to construct search and integration.," in *AMCIS*, p. 524, 2010.

[29] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, pp. 391–407, 1990.

[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*, pp. 3111–3119, 2013.

[31] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[32] V. K. Vaishnavi and W. Kuechler, *Design science research methods and patterns: Innovating information and communication technology*. New York, NY, USA: Auerbach, 1. ed., 2007.

[33] K. E. Weick, "Theory construction as disciplined imagination," *Academy of Management Review*, vol. 14, no. 4, pp. 516–531, 1989.

[34] M. Baroni, G. Dinu, and G. Kruszewski, "Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, pp. 238–247, 2014.

[35] M. Naili, A. H. Chaibi, and H. H. B. Ghezala, "Comparative study of word embedding methods in topic segmentation," *Procedia computer science*, vol. 112, pp. 340–349, 2017.

[36] C. D. Manning, C. D. Manning, and H. Schütze, *Foundations of statistical natural language processing*. MIT press, 1999.

[37] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.

[38] M. F. Porter, "Snowball: A language for stemming algorithms," 2001.

[39] S. T. Dumais, "Improving the retrieval of information from external sources," *Behavior Research Methods, Instruments, & Computers*, vol. 23, no. 2, pp. 229–236, 1991.

[40] P. Nakov, A. Popova, and P. Mateev, "Weight functions impact on lsa performance," *EuroConference RANLP*, pp. 187–193, 2001.

[41] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse processes*, vol. 25, no. 2-3, pp. 259–284, 1998.

[42] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[43] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of machine learning research*, vol. 3, no. Feb, pp. 1137–1155, 2003.

[44] O. Levy and Y. Goldberg, "Neural word embedding as implicit matrix factorization," in *Advances in neural information processing systems*, pp. 2177–2185, 2014.

[45] O. Levy, Y. Goldberg, and I. Dagan, "Improving distributional similarity with lessons learned from word embeddings," *Transactions of the Association for Computational Linguistics*, vol. 3, pp. 211–225, 2015.

[46] G.-I. Brokos, P. Malakasiotis, and I. Androutsopoulos, "Using centroids of word embeddings and word mover's distance for biomedical document retrieval in question answering," *arXiv preprint arXiv:1608.03905*, 2016.

[47] J. Raiman, "Python glove implementation." https://github.com/JonathanRaiman/glove, 2019.

[48] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, pp. 1188–1196, 2014.

[49] M. Chen, "Efficient vector representation for documents through corruption," *arXiv preprint arXiv:1707.02377*, 2017.

[50] M. Pagliardini, P. Gupta, and M. Jaggi, "Unsupervised learning of sentence embeddings using compositional n-gram features," *arXiv preprint arXiv:1703.02507*, 2017.

[51] Y. Li, D. McLean, Z. A. Bandar, K. Crockett, *et al.*, "Sentence similarity based on semantic nets and corpus statistics," *IEEE Transactions on Knowledge & Data Engineering*, no. 8, pp. 1138–1150, 2006.

[52] INN, "Inter-nomological network." https://inn.theorizeit.org/, 2019.

[53] D. Dann, A. Maedche, T. Teubner, B. Mueller, C. Meske, and B. Funk, "Toward a platform for the systematic accumulation of knowledge in is research," in *Proceedings of the 40. International Conference on Information Systems (ICIS 2019)*, 2019.

[54] P. R. Newsted, S. L. Huff, and M. C. Munro, "Survey instruments in information systems," *MIS Quarterly*, vol. 22, no. 4, pp. 553–554, 1998.