Cross-referencing using Fine-grained Topic Modeling

Jeffrey Lund, Piper Armstrong, Wilson Fearn, Stephen Cowley, Emily Hales, Kevin Seppi
Computer Science Department
Brigham Young University
{jefflund, piper.armstrong, wfearn,
scowley4, emilyhales, kseppi}@byu.edu

Abstract

Cross-referencing, which links passages of text to other related passages, can be a valuable study aid for facilitating comprehension of a text. However, cross-referencing requires first, a comprehensive thematic knowledge of the entire corpus, and second, a focused search through the corpus specifically to find such useful connections. Due to this, cross-reference resources are prohibitively expensive and exist only for the most well-studied texts (e.g. religious texts). We develop a topic-based system for automatically producing candidate cross-references which can be easily verified by human annotators. Our system utilizes fine-grained topic modeling with thousands of highly nuanced and specific topics to identify verse pairs which are topically related. We demonstrate that our system can be cost effective compared to having annotators acquire the expertise necessary to produce cross-reference resources unaided.

1 Introduction

Cross-references are references within or between bodies of text that can help elaborate upon or clarify that text. They can be a useful tool for deep understanding and can also be used to analyze the relational structure of a text. In contrast to a word concordance which simply shows passages which share a common keyword, cross-references often include links which do not necessarily share the same keywords, but are still related topically. The existence of a thorough and complete cross-reference resource can facilitate better scholarship of a text and help readers to quickly find clarifying information or see repeated themes throughout a text.

Compared to language tasks such as part-of-speech tagging, producing cross-reference annotations is much more labor intensive. To pro-duce cross-references, annotators must become intimately familiar with a text in order to note that a particular passage is related to another passage they happen to recall. This level of familiarity and expertise with a particular text typically requires the annotator to spend a great deal of time studying and reading. Possibly for this reason, expansive cross-references have only been produced for the most well-studied texts, such as the Bible. Other texts, such as academic textbooks, may include indices or other similar references, but these tend to be sparse, focusing on a small number of keywords rather than linking each individual passage with other relevant passages.

The process of creating such a resource can be expensive and time consuming. For example, the Bible[1] published by The Church of Jesus Christ of Latter-day Saints includes numerous cross-references and topic-based categories. These cross-references took hundreds of volunteers thousands of hours over seven years to produce (Anderson, 1979). This process involved collecting more than 19,900 manually curated entries from volunteers, and then editing and refining those references with a small committee of experts down to a final cross-reference database containing 12,475 entries.

Cross-referencing grows quadratically with the size of the data because for each of n passages, there are $n-1$ possibly related passages, yielding $\mathcal{O}(n^2)$ potential pairs. This differs from tasks such as part-of-speech tagging where annotators can tag individual sentences in isolation ($\mathcal{O}(n)$). We can, however, evaluate pairs in isolation. Therefore, our approach is to produce a system which utilizes fine-grained topic modeling in order to dramatically lower the cost of producing a cross-reference resource for new texts. We do not expect that such

---

[1]https://www.lds.org/scriptures/bible

3978

a system will produce only–or even primarily–valid cross-references, but we hope that the system could be accurate enough to allow annotators to simply review the proposed cross-references and reduce the search cost.

2 Methodology

In this section, we describe our experimental setup and how we approach the problem of automatic cross-reference generation.

2.1 Cross-reference Datasets

While exact statistics are impossible to obtain, the number of printed copies of the Bible is estimated to be more than 5 billion (Guiness). It is one of the most well-studied texts in existence, and one of the few texts in the world with extensive cross-reference data. We utilize the English Standard Version of the Holy Bible (Crossway, 2001) in order to validate our method. We use this specific translation of the Bible because it is used on `openbible.info`. While our work focuses on this specific religious text out of necessity, it can also be applied to other texts, including literary classics and collections of historical documents.

To aid any readers who are unfamiliar with this religious text, we note that the term 'verse' refers to a short division of a chapter. The entirety of the text is divided up into books. Typically, individual verses are referenced by the name of the book, the chapter number, and the verse number, e.g., Isaiah 25:8. For convenience, each referenced verse in electronic versions of this paper is a hyperlink to `openbible.info`, showing the verse along with associated cross-references. For the sake of narrowing our focus, our efforts in cross-referencing the Bible will focus on finding topically related verses, though other work could potentially link larger passages, such as entire chapters or passages spanning multiple verses.

As a ground truth for cross-references, we utilize two sources. The first is the "Treasury of Scripture Knowledge, Enhanced" (Morton, 2010) (an extended version of the original "Treasury of Scripture Knowledge" (Torry, 2002)), which includes 670,796 cross-references between the 31,085 verses of the Bible. To our knowledge, this is the most exhaustive resource of human curated cross-references for the Bible to date. We will denote this cross-reference dataset as *TSKE*.

The second source of ground truth cross-references is a dataset from `openbible.info`. This dataset was seeded with various public domain cross-reference data, including the Treasury of Scripture Knowledge. As shown in Figure 1, users search for a verse they are interested in and can then vote on whether they found a particular cross-reference to be helpful or not. With each helpful or not helpful vote counting as $+1$ and -1 respectively, the dataset includes the net result of the votes for each included cross-reference.

Thus we can filter the dataset of cross-references based on how helpful each verse was rated to be. Counting only those cross-references which have a non-negative vote total, this dataset contains 344,441 cross-references. In figures, we denote this subset of the `openbible.net` cross-reference dataset as *OpenBible+0*. We also use the subset of cross-references which received a net total of at least 5 helpful votes. This subset, denoted as *OpenBible+5*, has 50,098 cross-references.

We do note however that the voting data has some skewness in the number of votes for each cross-reference. The overwhelming majority of cross-references received fewer than five total votes for or against the reference. A small number of verses, including both popular verses as well as verses which happen to come from the very beginning of the Bible, have received hundreds of votes.

2.2 Baselines for Automated Cross-reference Generation

These cross-reference datasets were produced at a tremendous cost in time and human effort. To the best of our knowledge, efforts at automating this process are limited and have not received much attention in computer science literature. That said, a reasonable baseline for automated efforts is a simple word-based concordance, which lists words along with references to where the words occur in the text[2].

Using the TSKE as the ground truth for cross-references, this simple baseline will recover roughly 65% of the cross-references. For example, as shown in Figure 1 and assuming that stemming is performed, the verse Isaiah 25:8 would be properly linked to 1 Corinthians 15:54 due to the two verses sharing the terms 'death' and 'swallow'. On the other hand, verses such as Hosea 13:14 or 1 Corinthians 15:55 which reference the

---

[2]For an example of such a concordance for the King James Version of the Bible, see Strong's Concordance (Strong, 1890).

'sting of death' should not be linked to verses such as Revelation 9:10, which references the sting of a scorpion. For this reason, roughly 99% of cross-references found using word-based concordance are spurious according to the TSKE, making this baseline less useful as a cross-referencing resource.[3] We refer to this baseline as *word match*.

As a slightly stronger baseline, we also consider a topical concordance in which verses assigned to the same topic by some topic model are considered to be linked. We refer to this baseline as *topic match*. For example, suppose that a topic model includes a topic which gives high probability to terms such as 'death', 'swallow', 'victory' and 'sting'. Assuming that such a model would assign the previously mentioned verse to this topic, then verses such as Isaiah 25:8, Hosea 13:14, and 1 Corinthians 14:54 would be linked, but Revelation 9:10 which uses the term 'sting' in a different context (i.e., the sting of a scorpion) would not be linked.

We can further increase the precision of this baseline by only linking references which share a topic and a word, although this does come at the cost of recall. We refer to this final baseline method as *topic-word match*.

---

[3]For this reason, published biblical word concordances typically only give a manually curated subset of significant vocabulary terms.

2.3 Topic-based Cross-referencing

We now describe our approach to topic-based cross-referencing. The baselines built upon word or topic concordances simply propose any cross-reference for which a word or topic matches another verse, meaning that we cannot set a threshold on the quality of the proposed cross-references. Instead, we propose comparing document-topic distributions as K-dimensional vectors, where K is the number of topics, using standard vector distance metrics to compare verses. This idea has been used before (Towne et al., 2016), although not for the task of producing cross-references. By using a vector distance metric to compare the topical similarity of verse pairs, we can set a threshold on the number of proposed cross-references and propose only the most topically related verse pairs as cross-references. We experiment with four distance metrics: cosine distance, Euclidean distance, cityblock (or Manhattan) distance, and Chebyshev distance. However, given that previous work comparing document-topic vectors from LDA seem to default to cosine similarity (Towne et al., 2016; Chuang et al., 2012), we anticipate that cosine distance will be the best metric for selecting cross-references.

2.4 Model Selection

We claim that a fine-grained topic model, i.e., a topic model with a large number of highly nuanced topics, will be able to provide more value

Figure 1: Voting interface for cross-reference data from `openbible.info`.

for tasks like cross-referencing than traditional coarse-grained topic models. In order to validate this claim, we will compare our fine-grained models with topics from a traditional Latent Dirichlet Allocation model with 100 topics. We refer to this baseline model as *coarse*.

Traditional probabilistic topic models such as Latent Dirichlet Allocation are not able to utilize large numbers of topics (Wallach et al., 2009). However, we successfully train anchor-based topic models with thousands of topics. Consequently, for our fine-grained models, we will employ the Anchor Word algorithm (Arora et al., 2013). Anchor-based topic models view topic modeling as non-negative matrix factorization. This class of topic models attempts to decompose a document-word matrix into two matrices, including a topic-word matrix which gives the conditional probabilities of a particular word given a topic. Ordinarily, this factorization is NP-Hard (Arora et al., 2012). However, given a set of anchor words, or words which uniquely identify a topic, the computation requires only $\mathcal{O}(KV^2 + K^2VI)$ where K is the number of topics, V is the size of the vocabulary, and I (typically around 100) is the average number of iterations (Arora et al., 2013).

We train our anchor-based model using 3,000 topics. We choose this number based on the number of documents we expect each topic to explain: there are roughly 30,000 verses and, according to OpenBible+0, a median of 10 cross-references per verse, so we want each topic to be responsible for roughly 10 documents.

By default, the anchors for the 3,000 topics are produced using a modified form of the Gram-Schmidt process (Arora et al., 2013). This process views each word as a vector in high-dimensional space and attempts to pick anchor words which maximally span that space. For more details, see Arora et al. (2013). In our results and figures, we refer to this model with the default anchor selection method as *Gram-Schmidt*.

This does present us with some difficulty with anchor words as this process tends to select the most extreme and esoteric anchors possible (Lee and Mimno, 2014), which can lead to less useful topics as we increase their number. Lund et al. (2017) introduced a method of using multiple words to form a single anchor. This method, called tandem anchoring, was originally formulated as a way to extend the anchor algorithm to allow for interactive topic modeling.

Instead of utilizing human interaction to seed the topic anchors, we will seed the tandem anchors using the terms from randomly selected verses. For example, suppose we randomly select Isaiah 25:8 as a verse from which to form an anchor. As shown in Figure 1, this verse includes terms such as 'swallow', 'death', and 'tears'. Each of these terms is represented as a point in high-dimensional space. To produce a single anchor from these terms, we average the words using the element-wise harmonic mean[4]. While this new point may not correspond to any particular word, it does capture the joint occurrence pattern of the words which form the anchor. We repeat this process 3,000 times to produce an anchor-based topic model with tandem anchors. While this exact methodology of seeding topic anchors using randomly selected verses is novel, we note the similarity to the method used to seed topics in Rephil, a web scale topic model used by Google (Murphy, 2012). In figures, we refer to this model with tandem anchors as *tandem*.

For each of these models, we must take the topic-word distributions from the topic model and produce document specific topic assignments. We utilize mean field variational inference in order to assign the individual verses to topics, similar to Nguyen et al. (2015).

3 Results

In this section we present the results of our experiments with topic-based cross-referencing. As discussed in Section 2, we experiment with three different cross-reference datasets: TSKE, OpenBible+0, and OpenBible+5. We utilize three different topic models: coarse, Gram-Schmidt, and tandem. We seek to demonstrate that our topic-based cross-referencing system can effectively utilize fine-grained topic modeling to produce candidate cross-references which can be annotated by humans in a cost effective manner.

3.1 Metric Comparisons

We first explore the various metrics for selecting cross-references discussed in Section 2.3. With each proposed distance metric, we are able to set a threshold and determine which Bible verse pairs to keep as candidate cross-references and which

---

[4]See Lund et al. (2017) for details on why the harmonic mean is useful for forming tandem anchors

to discard. Figure 2 and Figure 3 summarize our results with respect to metrics.

Figure 2 gives a receiver operator characteristic (or ROC) curve, which compares the true positive rate (or recall) against the false positive rate (or fall-out). We show the curve for the various metrics using the TSKE as our ground truth. We show these curves on each of the three different models discussed in Section 2.4.

Overall, cosine distance is the best method for selecting cross-references, as it gives the largest area under the ROC curve. The major exception to this is with the traditional coarse-grained topic model for which Euclidean distance performs the best. Considering that cosine distance has frequently been used in conjunction with topics from LDA (Towne et al., 2016; Chuang et al., 2012), this result is somewhat surprising.

Also of interest is the fact that the word-match baseline does reasonably well with respect to the true positive rate, at least if a false positive rate of roughly 0.196 is acceptable. Note that this corresponds to 188,974,806 false positives in the TSKE dataset, so while the raw number of true positives may be impressive, this baseline is not likely to be useful in practice.

While the ROC plot is undoubtedly more popular than the precision-recall plot, in cases where the data is imbalanced, the precision-recall plot can be much more informative. This is mainly because of the use of false-positives in the ROC curve, which can present an overly optimistic picture of the classifier performance (Saito and Rehmsmeier, 2015). Consequently, in Figure 3, we also compare each of the proposed metrics using a precision-recall curve (or PRC).

The PRC plot reinforces the claim that cosine distance is the best distance metric to threshold cross-references since for any reasonable level of precision, cosine distance yields the best results. Once again, with the exception of Euclidean distance performing better with the coarse-grained topic model. However, we note that the precision using coarse-grained topics is much lower than using fine-grained topics with cosine distance. Even with Gram-Schmidt based fine-grained topics, where Euclidean distance eventually wins out against cosine distance, the high amount of false
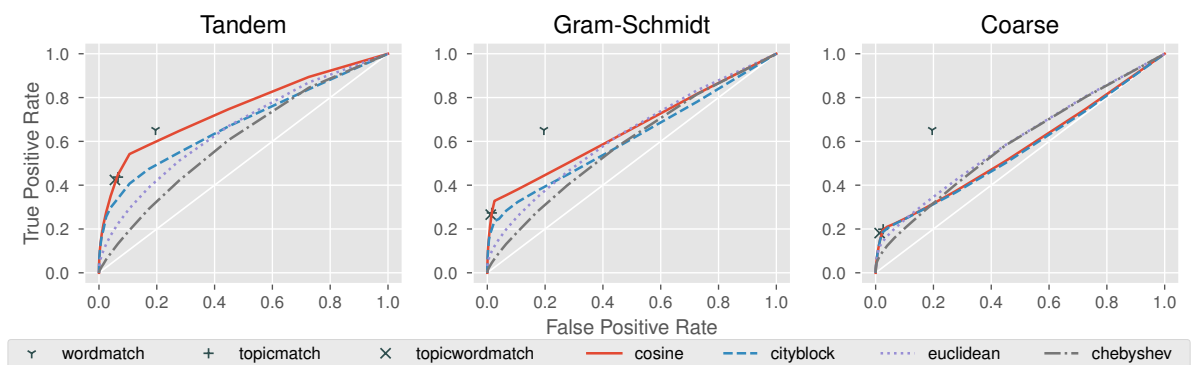

Figure 2: Cross-reference ROC curves for different metrics for cross-reference selection with topics from three different topic models and TSKE as the cross-reference ground truth.
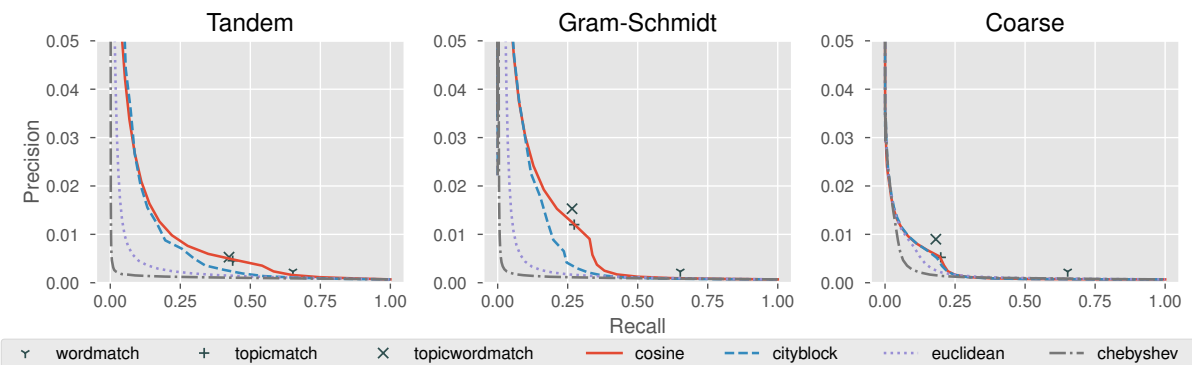

Figure 3: Cross-reference PRC plots with different metrics for cross-reference selection with topics from three different topic models and TSKE as the cross-reference ground truth.

3982

positives means that cosine distance is the most useful metric for this task. The PRC plot also illustrates why the matching baselines are not practically useful—they do have decent true positive rates, but the precision with these baselines is extremely low.

3.2 Topic Model Comparison

We now explore the various topic models discussed in Section 2.4. Figure 4 and Figure 5 summarize these results. Based on Section 3.1, each reported result in this section uses cosine distance to determine verse pairs which should be considered as candidate cross-references. We compare the results of the three topic models on the three datasets discussed in Section 2.1.

The overall trend on each of the three datasets is simply that we lower the true positive rate and precision as we use more selective datasets. Considering that OpenBible+5 is a subset of OpenBible+0, and that OpenBible+0 is nearly a subset

of TSKE[5], this result is not surprising. In the final cost analysis (see Section 3.3), the selectivity of the actual annotators must be taken into account when attempting to predict the true positive rate or the precision of a system.

As shown in Figure 4, regardless of the anchor selection strategy, both fine-grained topic models outperform the traditional coarse-grained model. Our model using tandem anchors built from randomly selected verses performs the best for nearly all levels of false positive rate. However, the Gram-Schmidt based anchors produce better true positive rates for very low false positive rates.

This trend is better illustrated with the PRC plots in Figure 5. While for higher values of recall the tandem anchor selection strategy does win out, it is only after precision significantly drops that tandem anchors produce superior predictions to Gram-Schmidt anchors.

---

[5]OpenBible+0 has 533 references seeded from other public domain sources, which are not included in TSKE.
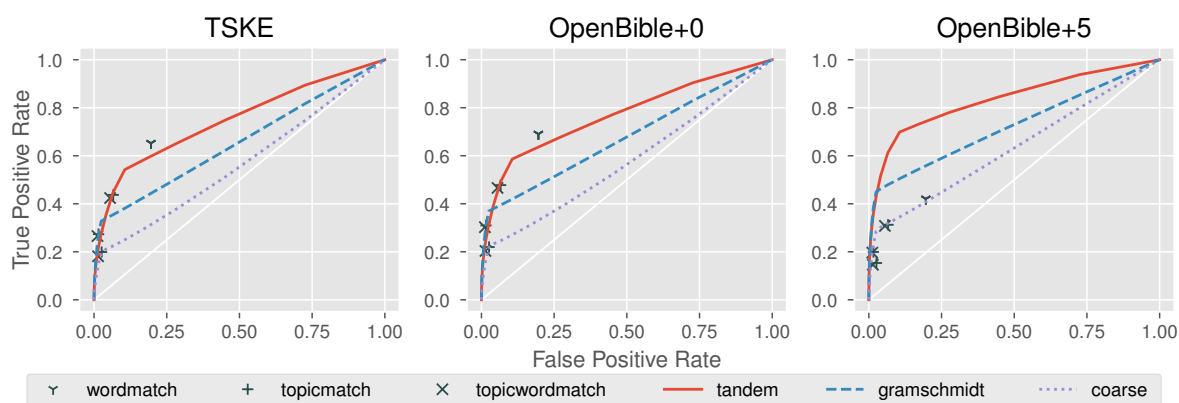

Figure 4: Cross-reference ROC curves with different models for cross-reference selection with three different datasets.
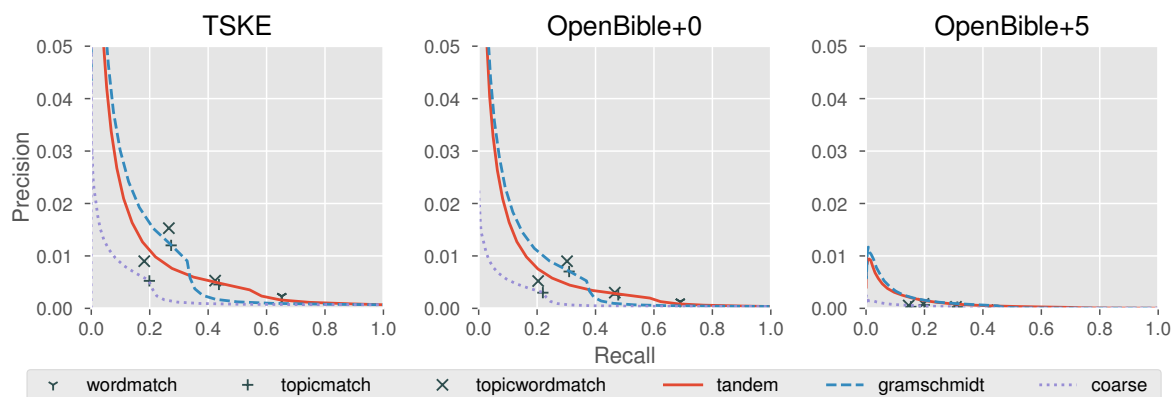

Figure 5: Cross-reference PRC plots with different models for cross-reference selection with three different datasets.
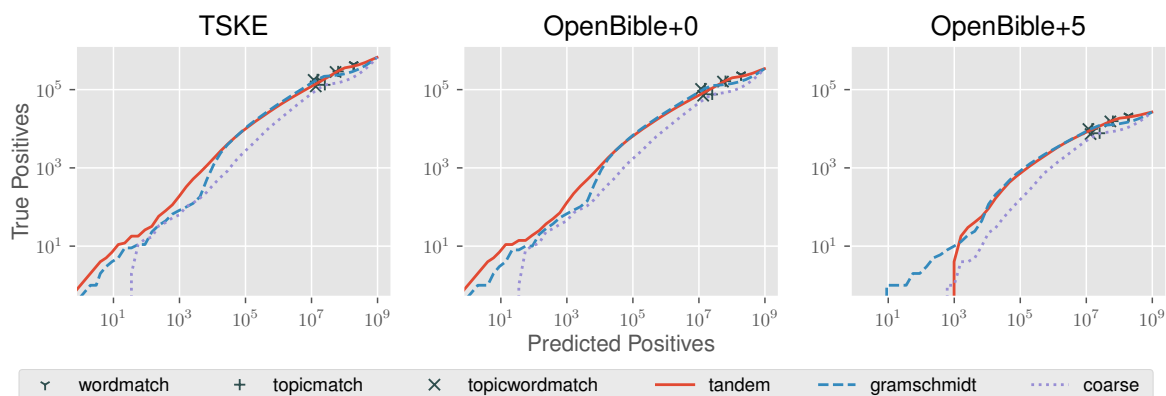
Figure 6: Cross-reference cost curves with different models for cross-reference selection with three different datasets. Note the log-log scale. The x-axis denotes the number of cross-references produced by our method, while the y-axis indicates how many of those cross-references are valid according to the human-provided ground truth.

3.3 Cost Analysis

While the precision-recall curves in Figure 5 may suggest that Gram-Schmidt based topic models produce superior topics for cross-referencing, we suggest that this analysis may be missing a key point in real world analysis. As an alternative to both PRC and ROC curves, we suggest that this task might be best served with an analysis of cost per true positive.

We envision that our system would be used to produce a set of candidate cross-references which would then be curated using human annotators. These annotators would be tasked with evaluating each potential reference and determining whether or not each cross-reference is valid. Critically, the annotator would only be required to evaluate individual cross-references, not the entire text.

As a working example of the cost of such an annotation process, suppose we use a popular crowdsourcing service (e.g., Amazon Mechanical Turk) to produce human annotations. We might reasonably expect to pay something around $0.01 USD per annotation. We would likely require some form of quality control in the form of redundant annotations, so we might end up paying $0.05 USD per annotated cross-reference candidate. Of course, the exact cost per annotated cross-reference will vary depending on the service and difficulty of the specific text being cross-referenced. However, we will use these estimates for the purpose of illustration.

Suppose as part of this working example, we are interested in producing a resource with 12,000 valid cross-reference annotations (roughly match-ing the size of the previously mentioned LDS edition of the Bible (Anderson, 1979)). Consulting Figure 6 we can then determine how many candidate cross-references we would need to produce for human annotation in order to create the final curated cross-reference resource. For example, using the TSKE as our ground truth, we would need approximately 150,000 predicted positives in order to find 12,000 true positives. At $0.05 USD per annotation, this would cost about $7,500 USD. Supposing that our annotators were more selective, we could use the OpenBible+0 as the ground truth, which would roughly double the cost. With OpenBible+5, which is considerably more selective, this cost rises to approximately $1,000,000 USD. In contrast, with traditional coarse-grained topic modeling, this cost is anywhere from $40,000 USD using the TSKE as the ground truth, to $17,500,000 USD using Open-Bible+5 as the ground truth.

While these costs may seem prohibitive, consider that the alternative is to have experts understand the entire text to the degree that they can read one passage and recall other relevant passages they have previously read. In the case of religious texts, this is often possible since adherents study those texts as part of their daily routine. For example, in the case of the LDS edition, it took a committee of experts seven years of work to produce their cross-reference resource, even with the aid of hundreds of volunteers. However, without those experts and volunteers, the cost would have been even greater. In the naive case where every possible reference is manually checked, the cost skyrockets to around $48,000,000 USD.

Passage 1	Passage 2
"I have an excessive regard for Miss Jane Bennet, she is really a very sweet girl, and I wish with all my heart she were well settled. But with such a father and mother, and such low connections, I am afraid there is no chance of it."	"I wish you may not get into a scrape, Harriet, whenever he does marry;–I mean, as to being acquainted with his wife–for though his sisters, from a superior education, are not to be altogether objected to, it does not follow that he might marry any body at all fit for you to notice. The misfortune of your birth ought to make you particularly careful as to your associates. There can be no doubt of your being a gentleman's daughter, and you must support your claim to that station by every thing within your own power, or there will be plenty of people who would take pleasure in degrading you."

Table 1: Passage 1 is taken from the eighth chapter of *Pride and Prejudice*, and Passage 2 is taken from the fourth chapter of *Emma*. These two passages are a valid cross-reference because they both discuss social standing and family connections in the context of marriage. Their connection was found even with their lack of shared words.

4 Discussion

Without extensive cross-referencing resources for more secular datasets, it is difficult to empirically prove the usefulness of our system generally without an extremely costly user study. That said, we make a small attempt by manually examining cross-references generated from the complete works of Jane Austen and Plato. Based on our cost analysis in Section 3.3, and since we will be examining only a small number of cross-references, we utilize tandem anchors to generate topics. With each dataset, we examine the first 300 cross-references produced by our system.

We also examine what our model got wrong in proposing Bible cross-references.

4.1 Jane Austen

Of the first 300 cross-references, we find that 39 of them are valid, linking passages from all six works by Jane Austen. As with our experiments with the Bible, this level of precision is sufficient that we believe that we could dramatically lower the cost of producing a full cross-referencing resource for this text.

We note that 109 of these are cross-references linking a paragraph in the eighth chapter of *Pride and Prejudice* to other passages in our corpus. Of the references involving this one paragraph, 22 were valid. An example of such a cross-reference is shown in Table 1. While marriage in general is a common theme in the works of Jane Austen, this particular paragraph more specifically discusses the role of social status and family connection as it relates to choosing a marriage partner. We note

that the connection between the passages in Table 1 is thematic; they share no significant words in common, demonstrating the capability of the system to detect nuanced topics and themes.

4.2 Plato

Of the first 300 cross-references generated from the works of Plato, we found 119 that were valid. Many of the cross references were between distinct works, and included discussions about the nature of justice, arguments about the composition of things, the nature and role that certain things play, and discussions of appropriate legislation.

It is important to note that the model found significantly more cross references in the works of Plato than those by Jane Austen. This is likely due to the nature of the writing. We find in the works of Plato that ideas themselves are discussed directly, similar to the bible, and thus we would expect it to be easier for a model to find words and phrases that link to a specific topic. An example of this is that words like "virtue," "courage," "mean," and "cowardice" would likely identify a topic about virtue that comes up in the works of Plato. However, in Jane Austen we find ideas discussed implicitly through interactions of the characters and commentary by the author. The meaning, while present, is found by "reading between the lines." An example of this is that we might find marriage discussed in the works of Jane Austen, but more often through characters discussing their feelings after getting married. Here, words that we might see could be words such as "happy," "elated," "love," "efficient," etc. However, these

words could also correlate equally well with other topics, and thus it would be harder for our model to discern. As further evidence of this we point out that the cross reference we used above from Jane Austen is an explicit discussion of marriage. It is likely that an implicit discussion of marriage would be harder for our model to find. We also point out that in such cases it is a non-trivial task for humans to come to a specific consensus about what a given passage could mean or relate to.

That said, given the relevant and influential nature that the works of Plato still hold even today, we can see that these cross references are highly useful in that they facilitate study and understanding of his works that a study of each individual work separately might miss.

4.3 Error Analysis

We examine the errors of running tandem anchors using cosine distance on the Bible. There are two types of errors to examine: candidate cross-references proposed early in the process that are not valid cross-references and valid cross-references that are not proposed until the end of the process (as determined by the Treasury of Scripture Knowledge).

Early invalid candidate cross-references all exhibit the same characteristic; the documents are exactly or substantively the same (e.g. *Deuteronomy 2:17 the Lord said to me,* and *Deuteronomy 2:2 Then the Lord said to me*). Indeed, a human given only those two documents would also mark them as related, and many valid cross-references exhibit this same characteristic (e.g. Psalms 107:6 and Psalms 107:28 are exactly the same and are a valid cross-reference).

Cross-references are partially so difficult because what constitutes a valid cross-reference is at least partially determined by what the community surrounding the text views as significant, and so two documents that are identical may or may not be a valid cross-reference depending on the view of that community. This particular issue would make any end-to-end automated solution to cross-referencing particularly difficult.

We also examined the last one hundred valid cross-references proposed. We consider these errors because in a system where a predetermined number of candidate cross-references are considered, these candidate cross-references would most likely never be considered. For 28 of the last one

hundred valid cross-references, it is unclear why they are considered valid cross references (e.g., *Ezekiel 16:10 I clothed you also with embroidered cloth and shod you with fine leather. I wrapped you in fine linen and covered you with silk.* and *Deuteronomy 8:11 "Take care lest you forget the Lord your God by not keeping his commandments and his rules and his statutes, which I command you today,*).

Many of the other 72 valid cross-references are also difficult in some way. Many of the connections involve some use of metaphor (25) or are linked by a single key word, such as a name, but are otherwise topically dissimilar (34). The other 13 cross-references are all documents describing the construction of the tabernacle, and we have enough extra context to recognize this, however it isn't surprising that the model doesn't find this connection.

It is useful to note that of the final 600,000 candidate cross-references, only one hundred of them were valid cross-references.

5 Conclusion

We have produced a system using fine-grained topic modeling which is able to propose candidate cross-references which can be verified by non-expert human annotators for the purpose of creating a cross-reference resource at a fraction of the cost of current manual techniques. Our method, which utilizes tandem anchors to produce large numbers of highly nuanced topics coupled with an effective assignment strategy, is able to produce document-topic vectors which are comparable using cosine distance.

Our results also demonstrate that this system would not be as cost effective with traditional coarse-grained topic modeling. While we can find sets of topically related documents using coarse-grained topics, for the task of finding the most closely related documents we require a system which is more specific. We suggest that this success serves as motivation for exploration of fine-grained topic modeling for other topic-based use cases which require nuance and precision.

Acknowledgements

References

Lavina Fielding Anderson. 1979. Church publishes first LDS edition of the bible. *Ensign*, 10.

Sanjeev Arora, Rong Ge, Yonatan Halpern, David Mimno, Ankur Moitra, David Sontag, Yichen Wu, and Michael Zhu. 2013. A practical algorithm for topic modeling with provable guarantees. In *Proceedings of the International Conference of Machine Learning*.

Sanjeev Arora, Rong Ge, and Ankur Moitra. 2012. Learning topic models–going beyond svd. In *Fifty-Third IEEE Annual Symposium on Foundations of Computer Science*.

Jason Chuang, Daniel Ramage, Christopher Manning, and Jeffrey Heer. 2012. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 443–452. ACM.

Crossway. 2001. *The Holy Bible: English Standard Version*. Crossway Bibles.

Guiness. 2018. Best selling book of non-fiction. Http://www.guinnessworldrecords.com/world-records/best-selling-book-of-non-fiction.

Moontae Lee and David Mimno. 2014. Low-dimensional embeddings for interpretable anchor-based topic inference. In *Proceedings of Empirical Methods in Natural Language Processing*.

Jeffrey Lund, Connor Cook, Kevin Seppi, and Jordan Boyd-Graber. 2017. Tandem anchoring: A multi-word anchor approach for interactive topic modeling. In *Proceedings of the Association for Computational Linguistics*.

Timothy Morton. 2010. *Treasury of Scripture Knowledge, Enhanced*. BibleAnalyzer.com.

Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*, pages 928–929. MIT Press.

Thang Nguyen, Jordan Boyd-Graber, Jeffrey Lund, Kevin Seppi, and Eric Ringger. 2015. Is your anchor going up or down? Fast and accurate supervised topic models. In *Conference of the North American Chapter of the Association for Computational Linguistics*.

Takaya Saito and Marc Rehmsmeier. 2015. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432.

James Strong. 1890. *The Exhaustive Concordance of the Bible*. Jennings & Graham.

R.A. Torry. 2002. *Treasury of Scripture Knowledge*. Hendrickson Publishers.

W Ben Towne, Carolyn P Rosé, and James D Herbsleb. 2016. Measuring similarity similarly: Lda and human perception. *ACM Transactions on Intelligent Systems and Technology*, 8(1).

Hanna M Wallach, David M Mimno, and Andrew McCallum. 2009. Rethinking lda: Why priors matter. In *NIPS*, volume 22, pages 1973–1981.