Modeling Influence in Text Corpora

Anonymous Author(s)
Affiliation
Address
email

Abstract

Identifying the most influential documents in a corpus is an important problem in many fields, ranging from information science and historiography to text summarization and news aggregation. We propose using changes in the linguistic content of these documents over time to predict the importance of individual documents within the collection and describe a dynamic topic model for both quantifying and qualifying the impact of each document in the corpus.

1 Introduction

Computing meaningful statistical measurements of the influence of a scientific article is an important and challenging problem. Such measurements are used to assess the quality of various academic instruments, such as journals, scientists, and universities. These so-called ”impact factors” can play a large role in decisions surrounding funding, hiring, and publishing.

Knowing the impact of a scientific article is important for scientific research as well: finding and reading the influential articles of a field is central to good research practice. With unprecedented access to large corpora of scientific articles, modern researchers have instant access to a large catalog of previous work that might be important and interesting.

The traditional method of assessing the impact of an article is to analyze the citations to it. The impact factor of a journal, for example, is based largely on citation analysis; and this is intuitive: if more people have cited an article then more people have read it. It has had more impact on its field.

The notion of impact or influence goes beyond the small world of scientific research, however: legal documents, news stories, blog posts, sermons, and email all might have had an impact on other documents. But, how can we measure impact without citations? Moreover, citations only capture one kind of impact. All citations from an article are counted equally in an impact factor. However, some articles of a bibliography might be more related–or have influenced the authors more–than others.

The goal of this work is to develop an unsupervised method for determining the influence of a document. Our intuition is that language changes over time, and that influential documents contribute to this change. Our algorithm takes a sequence of documents as input and computes a scalar ”influence” of each document. We validate our method by measuring how well the computed impact predicts citation. On a corpus of over 20 years of Computational Linguistics articles, we demonstrate that our method provides a citation-free measure of bibliometric impact.

This paper is organized as follows. We begin with a discussion of work related to this . We then describe the Document Influence Model (DIM), our unsupervised model for determining the influence of a document using the changes in language used by documents over time. We then provide details for inference with this model. We then discuss experimental results and provide examples of documents deemed influential by this model.

2 Related work

Much of the work in identifying influential documents has focused on additional forms of structure within a corpus such as hyperlinks or document citations. The utility of citation counts in gauging influence is a well-established in practice, supported by a wide body of research and practical applications. The *Impact Factor*, for example, a measure of the importance of a journal within a field of research, is based on citations to this journal's articles [1]. The well-known Pagerank algorithm has likewise helped Google to find documents relevant to Web search queries [2]. In these cases, this structure has been very powerful: recieving a Nobel prize is highly correlated with citations to one's work (before the prize is awarded), for example; and Google has become one of the world's foremost search engines.

Much recent work has gone into understanding the interplay between corpus semantics and this link structure. Radev et al. have studied the relationship between semantic content of Web documents and the probability a hyperlink existing between them. Deitz et al. have used topic models, a formalism for identifying themes in corpora, to characterize the the thematic significance of individual citations between documents [3].

2.1 Topic models

We will also build on *topic models*, a formalism for understanding themes in corpora. Examples of topic models include Latent Semantic Indexing (LSI) [4], probabilistic LSI (pLSI) [5], and Latent Dirichlet Allocation (LDA) [6]. In each of these models, each document is modeled as a mixture of *topics*, which can generally be interpreted as themes.

Existing work has aimed to explicitly predict these links using themes in a corpus. Cohn et al. have used topic models for inferring links (both hyperlinks and academic citations) between documents [7]. Relational Topic Models and Link-LDA have also been posited to model citations and hyperlink references in the Blogosphere, respectively [8, 9].

While this work is related to our goal, we do not aim to predict individual links between documents; we are primarily interested in understanding which documents are influential, and *how* these documents are influential. While the model in [7] can be interpreted as a stochastic process which mimics the generation of documents, we explicitly aim to use the evolution of topics *over time* to determine influential documents.

Various applications of topic models have addressed the evolution of themes over time. *Post-hoc* analyses have been performed to track the progression of ideas over time, identifying "hot" and "cold" topics in the Journal PNAS [10] and the ACL Anthology [11]. More direct methods for understanding the flow of ideas over time involve extensions of topic models to capture the temporal evolution of themes. The Dynamic Topic Model [12], makes a Markov assumption about how topics drift over discrete points in time and has been extended to a continuous model [13]. Topics over Time (TOT) [14] models topics as fixed distributions whose proportions over time are given by beta distributions, an assumption convenient for inference. The Dynamic Mixture Model (DMM) [15], effectively an online model, assumes a fixed set of topics but allows the topic proportion to vary over time. In contrast to the DTM, the DMM assumes that topics themselves do not evolve over time; only the (prior over) topic proportions change over time.

2.2 Dynamic Topic Models

In this work, we assume that influential documents change the language of their topics. The DTM is therefore a natural basis on which to build our model, the Document Influence Model. Before describing the DIM in detail, we give a quick review of LDA and the DTM; readers can refer for further details to the original work in [6] and [12]. In LDA, a corpus is associated with a parameter α describing a Dirichlet prior over topic proportions assigned to each document, along with a set β of topics which describe themes as distributions over words.

In the DTM, documents at *each point in time* are associated with distinct α_t (now as a logistic normal distribution) and topics β_t; within each time point, documents are assumed to be generated by the LDA generative process. The DTM further assumes that these topics and topic proportions drift over time. In particular, the natural parameters $\beta_t = \log(\pi_t)$ of each topic drift as a Gaussian process with fixed chain variance σ^2.

3 Document Influence Model

The Document Influence Model (DIM) identifies documents which change the focus of discussion in subsequent documents, i.e., those documents which change the language used by future documents.

To do this, we extend the dynamic topic model by associating each document d_t with a set of topic weights $\vec{l}_{d_t,k}$ which express how much the language used in document d_t affects the natural topic parameter β, and hence the language, used by subsequent documents. The more influential a document is, the larger its topic weights will be, making it more likely that future documents will use the same language.

The influence of a document works as follows. First, consider a topic $\beta_{t,k}$. We assume that documents about topic $\beta_{t,k}$ also have some influence on the language used within the topic. More specifically, we assume that they linearly "nudge" its natural parameters based on the words in these documents. Because topic β_k is not likely to be affected by documents or words unrelated to it, we only allow topic $\beta_{t,k}$ to be affected by words already assigned to it. In LDA, this topic assignment for word n is given by its the random variable $\mathbf{z}_n$.

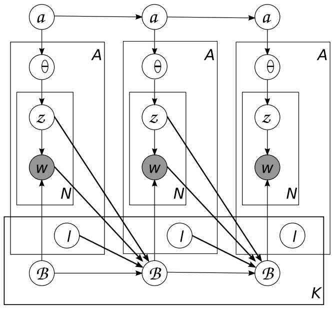

Figure 1: The Document Influence Model.

The full generative model at time t is then:

1. For topic $k = 1, \ldots, K$:
   Draw $\beta_{t,k} | \beta_{t-f,k}, \mathbf{z}_{t-f}, l_{t-f} \sim \mathcal{N}(\beta_{t-f,k} + \frac{\partial \beta}{\partial t}, \sigma^2 I)$.
2. For each document d_t:
   (a) Generate all documents at time t using LDA with parameters α_t and β_t.
   (b) For topic $k = 1, \ldots, K$, draw document weight $l_{d,k} \sim \mathcal{N}(\mathbf{0}, \sigma_l^2 I)$.

Here $\frac{\partial \beta}{\partial t} = \frac{\partial \exp \beta_{t-f,k}}{\partial t} \exp(-\beta_{t-f,k}) = \exp(-\beta_{t-f,k}) \circ (([\mathbf{z}_{t-f}]_k \circ \mathbf{W}_{t-f})(l_{t-f,k}))$, with $\circ$ denoting the element-wise Hadamard product, f the delay before a document's effect takes place, and $[z]_k$ the indicator describing whether term z is in topic k. Figure 1 depicts this generative process as a graphical model.

With the generative model described above, we can express the likelihood of the data as

$$p(\mathbf{d}_{1:T}) = \prod_K \prod_T p(\mathbf{W}_t | \beta_t, \mathbf{z}_t) p(\beta_t | \beta_{t-f}, \mathbf{z}_{t-f}, l_{t-f}, \mathbf{W}_{t-f}) \prod_{D_t} p(l_{d,k}). \tag{1}$$

4 Inference and parameter estimation

With the model defined, the desiderata are to infer the latent variables and estimate the hidden parameters of the model given observed data. Determining the distribution over latent variables is known as posterior inference. This is intractable to compute exactly, so we employ mean-field variational approximate inference. We couple this with variational EM to estimate the hidden parameters of the model. See Jordan et al. [16] for a review.

Variational posterior Variational inference in the Dynamic Topic Model is accomplished by structured variational inference: inference within each point of time t is treated as LDA [6] for topics β_t, and the sequential structure of the topics is retained with a collection of variational parameters $\tilde{\beta}_1, \ldots, \tilde{\beta}_T$. This structured approach makes inference tractable while keeping the variational posterior expressive and flexible [12]. We add an additional variational parameter $\tilde{l}_{d,k}$ for each (document × topic). These parameters fully factorize with the rest of the posterior, each yielding the marginal posterior $l_{d,k}|\tilde{l}_{d,k} \sim \mathcal{N}\left(\tilde{l}_{d,k}, \sigma_l^2\right)$ [1]. The full variational posterior is then

$$\begin{aligned} q(\mathbf{d}) &= \prod_{k=1}^{K} q(\beta_{k,1}, \ldots, \beta_{k,T}|\tilde{\beta}_{k,1}, \ldots, \tilde{\beta}_{k,T}) & (2) \\ &\times \prod_{t=1}^{T} \left(\prod_{d=1}^{D_t} q(\theta_{t,d}|\gamma_{t,d}) q(l_{t,d}|\tilde{l}_{t,d}) \prod_{n=1}^{N_{t,d}} q(\mathbf{z}_{t,d,n}|\phi_{t,d,n}) \right) & (3) \end{aligned}$$

The evidence can be bounded from below using Jenson's inequality. Maximizing this lower bound corresponds to minimizing the variational distribution's KL divergence with the true posterior. This lower bound and details of optimizing it are provided in the appendix.

While inference for this model is similar to that described in [12], some updates are new or significantly different. Here we focus on the update equation for the new parameter $\tilde{l}$ and changes to the update equations for ϕ, the variational parameter for z, and the new parameters $\tilde{l}$. We provide a derivation of the update for β in an appendix, and readers can refer to [12] for further details on optimizing this parameter.

Document regression As suggested by the generative model given in section 3, the update for $\tilde{l}$ is the solution to a regression problem. This solution is:

$$\begin{aligned} \tilde{l}_{t,k} \quad &\leftarrow \Big(\tfrac{\sigma^2}{\sigma_d^2} I + \Big((\mathbf{W}_{t,k} \circ \phi_{t,k})^T \Lambda_{\exp(-2\tilde{m}+2\tilde{V})} (\mathbf{W}_{t,k} \circ \phi_{t,k}) \Big) & (4) \\ &\qquad + \Lambda_{(\mathbf{W}_{t,k} \circ \mathbf{W}_{t,k} \circ (\phi_{t,k} - \phi_{t,k} \circ \phi_{t,k}))^T \exp(-2\tilde{m}+2\tilde{V})} \Big)^{-1} \\ &(\mathbf{W}_{t-f,k} \circ \phi)^T \left(\exp(-\tilde{m}_{t,k} + \tilde{V}_{t,k}/2) \circ (\tilde{m}_{t+f,k} - \tilde{m}_{t,k} + \tilde{V}_{t,k}) \right), & (5) \end{aligned}$$

where $\circ$ refers to the Hadamard element-wise product and $\Lambda_{\vec{x}}$ refers to a diagonal matrix having the elements of $\vec{x}$ on its diagonal. $\mathbf{W}$ here is the term-document matrix, and $\tilde{m}$ and $\tilde{V}$ are the variational expectation and variance of β, as described in [12].

This update provides some insight into how the model works: it assumes that document weights are generated by a regularized linear regression with a random design matrix $X := \mathbf{W} \circ \phi$. This update has some very nice properties. First, regularization is enforced by three separate forces: the prior over document weights, the variance σ^2 of topic drifts, and the variance $\phi(1-\phi)$ of the topics assigned to each word; we address the importance of this regularization in Section **??**. Because the topic distribution for each document is nearly sparse, with suitable transformations of W (such as column normalization), the document weights for a given topic can also be made sparse (in the sense that most weights are close to 0).

In the variational implementation of LDA [6], the ϕ updates are accomplished in a single pass. In the Document Influence Model, we cannot fully optimize ϕ in a single pass, but we can approximate it in each pass using an additional parameter λ_n and the principal branch W_0 of the Lambert function

[1] In general, we may abuse notation and refer to the vector $\tilde{l}_{t,k}$ of all document weights at time t for topic k; notation should be evident from context.

[2] as follows [17]:

$$\begin{aligned}
A_k :=& \frac{2}{\sigma^2}\exp(-2\tilde{m}_{t,n,k} + 2\tilde{V}_{t,n,k})(\tilde{l}^2_{t,k,d_n} w^2_{d_n}) \\
B_k :=& \Psi(\gamma_i) - \Psi(\sum_{j=1}^{K}\gamma_j) \\
& + \tilde{m}_{t,n,k} - 1 + \lambda_{n,s} \\
& + \frac{1}{\sigma^2}(\tilde{m}_{t+f,n,k} - \tilde{m}_{t,n,k} + \tilde{V}_{t,n,k})\exp(-\tilde{m}_{t,n,k} + \tilde{V}_{t,n,k}/2)\tilde{l}_{t,k,d_n} w_{t,d,n} \\
& - \frac{1}{\sigma^2}\exp(-2\tilde{m}_{t,n,k} + 2\tilde{V}_{t,n,k})(\tilde{l}_{t,k,d_n} w_{t,d,n} D_{n,k}) \\
& - \frac{1}{2\sigma^2}\exp(-2\tilde{m}_{t,n,k} + 2\tilde{V}_{t,n,k}) w^2_{t,d,n}(\tilde{l}^2_{t,k,d_n} + \vec{\sigma}^2_{l\ D_t}) \\
\phi_{n,k} \leftarrow & - \frac{W(-A_k e^{B_k})}{A_k}.
\end{aligned} \tag{6}$$

Here $D_{n,k} := \sum_d \phi_{d,n,k}\tilde{l}_{t,d,k} w_{t,n} := \Big((\mathbf{W}_t \circ \phi_{t,k})\tilde{l}_{t,k}\Big)_n$ can be stored with each update of $\tilde{l}$. Here A_k is the coefficient of ϕ, while B_k is the constant term (detailed in supplementary material). When A_k and $\sigma^2_{lD_t}$ are small (as is usually the case), this formula gives the same result as that found by LDA.

We then normalize $\phi_{n,k}$ over the topics k and update λ_n:

$$\lambda_{n,s+f} \leftarrow - \frac{\sum_K A_k\phi_k + B_k - \log(\phi_{n,k})}{K} + \lambda_{n,s}.$$

For further details on the derivation of this update, please see supplementary material.

The update for $\tilde{\beta}$ is similar to the update in [12] and can be determined from the evidence lower bound. We provide further details in the Appendix.

5 Experiments

We have designed the Document Influence Model to identify highly influential documents using only the changes in language over time. To confirm that this model is finding patterns of influence, we have validated it on a corpus in which documents have been annotated with internal citation counts. In this section we describe two experiments designed to understand whether there exists a relationship to existing influence metrics and to understand how to set the model's parameters. We also give some intuition into the model by showing papers deemed influential by it and illustrating two documents in detail.

5.1 Methodology

Predicting citations The principal experiment aims to determine whether the influence score is correlated with citation counts. We fit a model over a standard corpus using each document's true date. We perform this experiment with $\sigma^2_l = 0.00001$, letting $f = 2, 6,$ or 10. For simplicity, we restrict to a single topic, although we have successfully run this model with multiple topics.

Varying σ^2_d We also ran a set of experiments to understand when this model best matches citation data. Here we varid σ^2_d, the variance of document weights, for a one-topic model, selecting a range of values between 1e-10 and 5e-2. Changing σ^2_d amounts to changing regularization in Formula 4: smaller σ^2_d yields higher regularization.

5.2 Data

The *ACL Anthology* is a collection of papers spanning several decades in the field of computational linguistics. Joseph and Radev et al. have constructed the ACL Anthology Network from this corpus, annotating references and collecting various network statistics including paper citation counts,

[2] The Lambert function $W(y)$ is the solution to $xe^x = y$ and is implemented in GSL.

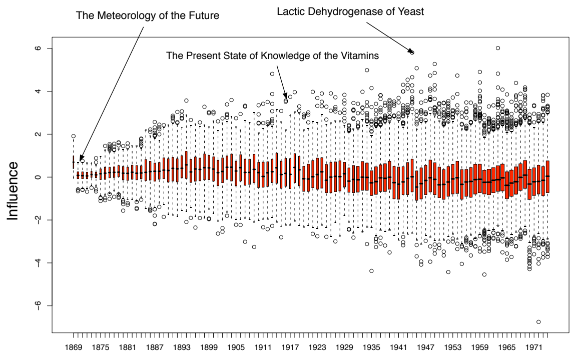

Figure 2: Document weights as a function of date. Titles of outliers are labeled.

author citation counts, and Pagerank [18] [19]. We fit the DIM on the all documents in the ACL Anthology Network for which dates were available. This included 11115 documents, from 1965 to 2007, discretized into two-year periods. To focus on high-content words, we restricted the corpus to plural nouns and common singular nouns using the TreeTagger part-of-speech tagger [20], for a total word count of 10,718,210 and a vocabulary of size 4,148.

To illustrate the DIM in a corpus where there is no citation information, we fit a one-topic DIM to the full text of Nature magazine from 1869 to the present. The box plot of posterior influences is illustrated, with some of the high impact articles identified. The DIM provides a new exploratory tool for scientific corpora, like this one, for which bibliometric information is difficult to obtain.

5.3 Results

We verified the relationship between the posterior influence and a notion of bibliometric impact. We fit linear regression models for each year of the ACL corpus of the log of citations (plus one) against the posterior influence. We plot the coefficient in Figure 3. To test its significance, we shuffled the impact factors 100 times and reran the regression, plotted in gray. (This is a permutation test.)

Figure 2 indicates that there is a definitive relationship between our estimate of influence and the number of citations of a document. We emphasize that our estimate of impact is based only on the way language has changed. From this information alone, we were able to infer a quantity that is predictive of bibliometric impact.

We also fit a 20 topic DIM model to the ACL data. Here we found a more significant relationship between posterior influence and citation. Specifically, the single-topic model acheived an adjusted R2 of 0.07, i.e., it captured 7The 20 topic model achieved an adjusted R2 of 0.18. This is a marked improvement. Moreover, we can investigate which topics contributed to the signal for further exploratory analysis.

5.4 Example influential documents

What are the most influential documents found by the DIM? In the supplementary material we provides a list of documents deemed most influential by the DIM in the ACL Anthology, along with their median citation counts, and Figure 4 shows a distribution of scores, with several outliers labeled. Here the ACL Anthology was prepared per Section 5.2 with σ_d^2 = 0.00001 and $f = 2$.

Measuring the influence of terms To better understand the model, and the regression in Equation 4 in particular, we can see which terms are "pulling" on a specific document. We define the "residual improvement" (RI) as follows. Given a linear regression $X\beta = Y + E$, some covariate index c, and a response index r, the RI is given by $(X_{r,-c}\beta_{-c} - Y_r)^2 - E_r^2$, where the subscript $-c$ denotes omission of a column or row from these respective vectors. This metric provides an indication of the significance of a covariate in the prediction of a particular feature. In our case, we omit documents to determine their imact on prediction of certain terms changing within the topic. This can be interpreted as the "pull" of individual terms on documents; inversely, it can be interpreted

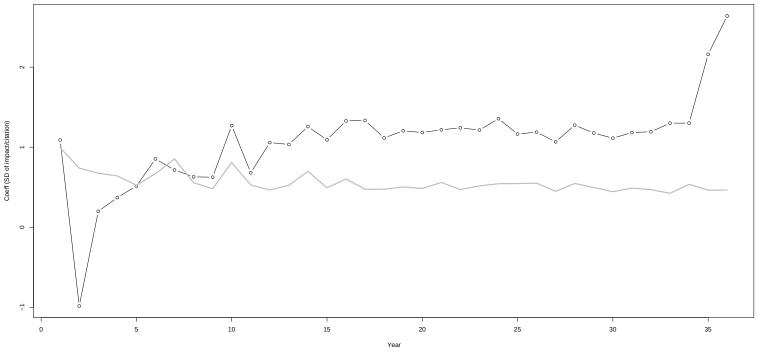

Figure 3: Coefficients of posterior influence regressed on log citations.

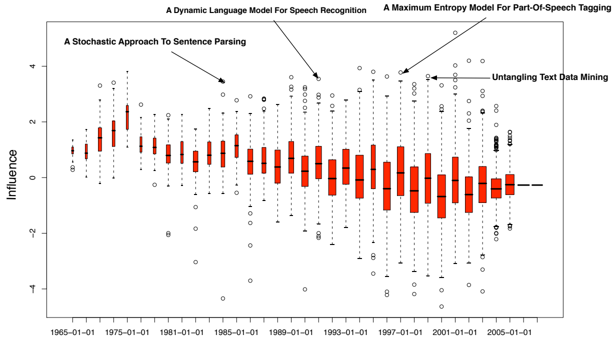

Figure 4: Document weights as a function of date. Titles of outliers are labeled.

as the direction a document helps to "push" a term. We can illustrate this metric with documents having scores at both the high and low extremes in the ACL Anthology.

The article *A Maximum Entropy Model For Part-Of-Speech Tagging* was the highest-influence document published in 1996. It had 215 citations in the ACL Anthology. Figure 5 shows this document in greater detail, illustrating a collection of its terms which most closely match the language of the corpus (i.e., have high values in $\mathbf{W} \circ \phi$). The most "significant" term (by RI) that this document is contributing to the language of the topic, for example, is "word"; following this are "model", "feature", and "datum", terms suggestive of the shift toward classification and probabalistic methods in the language of the corpus. The *least* significant term, by contrast, is "identity"; its negative weight, however, is trivial.

An Ascription-Based Approch To Speech Acts was the article in the ACL Anthology deemed *least* influential by the model in 1996. In contrast to *A Maximum Entropy Model*, it had zero citations. Figure 5 provides a closer view of this document, illustrating both the dynamics of this document's vocabulary over time and the terms towards which this document most contributes. Fits for the terms "speech" and "goal" were helped by this document, because its negative score corresponds to these words decreasing when the document appeared. The fit for "example", however, was hurt by this document; conversely, it pulled this document in the positive direction.

6 Discussion and future work

We have presented a model for identifying candidate influential documents, including details of a variational implementation and experiment results. While this model is modest, its good performance in identifying influential documents based solely on the changes in a corpus's language is a promising start.

A Maximum Entropy Model For Part-Of-Speech Tagging **Score: 3.63**	*An Ascription-Based Approach To Speech Acts* **Score: -4.22**

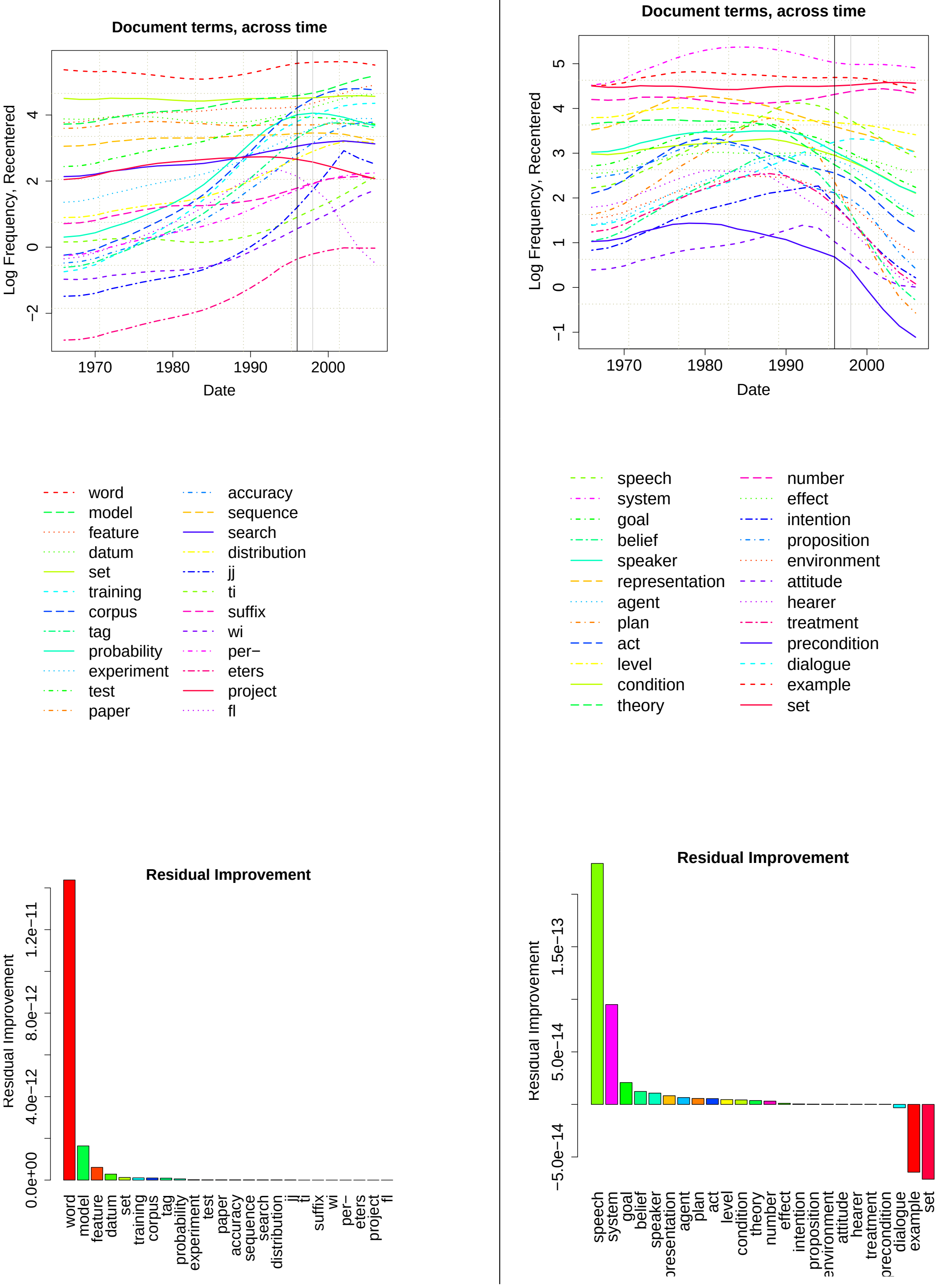

Figure 5: Selected documents from the ACL Anthology having highest and lowest scores in 1996.

In future work we hope to experiment with several variants of this model. We suspect, for example, that a model in which the influence of a document is smeared out over a period of time – through an *envelope* of influence, instead of being applied instantaneously – may be a more realistic model of the influence of a document. We further expect that using a nonnegative prior over the influence of a document will yield a more realistic model of a document's influence and minimize *post-hoc* transformation of scores such as capping. In addition, we hope to explore a richer set of models of the flow of ideas, including models of topics as birth and death processes (also noted in [12]) and tracking which documents may influence other documents.

We also hope to better understand this model in the context of traditional metrics of influence, such as academic citations, when such metrics are available. Understanding when this model and citation counts differ, for example, will help us to understand where our metric may fall short when applied to corpora in the wild. Furthermore, such comparisons will provide insight not just into this model, but also possibly into shortcomings or assumptions of these well-established metrics.

References

[1] Garfield, E. "Algorithmic citation-linked historiography - mapping the literature of science." *ASIST 2002: Information, Connections, and Community*, 2002.

[2] Brin, S., L. Page. "The anatomy of a large-scale hypertextual web search engine." In "Computer Networks and ISDN Systems," 107–117. 1998.

[3] Dietz, L., S. Bickel, T. Scheffer. "Unsupervised prediction of citation influences." In "ICML," 2007.

[4] Deerwester, S., S. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman. "Indexing by latent semantic analysis." *Journal of the American Society for Information Science*, 1990.

[5] Hofmann, T. "Probabilistic latent semantic analysis." *Proceedings of the Twenty-Second Annual International SIGIR Converence on Research and Development in Information Retrieval*, 1999.

[6] Blei, D. M., A. Y. Ng, M. I. Jordan. "Latent dirichlet allocation." *Journal of Machine Learning Research*, 2003.

[7] Cohn, D., T. Hofmann. "The missing link - a probabilistic model of document content and hypertext connectivity.", 2001.

[8] Chang, J., D. M. Blei. "Relational topic models for document networks." *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AIStats) 2009*, 5, 2009.

[9] Nallapati, R., W. Cohen. "Link-plsa-lda: A new unsupervised model for topics and influence of blogs." *International Conference for Weblogs and Social Media*, 2008.

[10] Griffiths, T. L., M. Steyvers. "Finding scientific topics." *Proceedings of the National Academy of Sciences*, 5528–5235, 2004.

[11] Hall, D., D. Jurafsky, C. Manning. "Studying the history if ideas using topic models." *Proceedings of EMNLP KW*, 2008.

[12] Blei, D., J. Lafferty. "Dynamic topic models." *Proc. of the 23rd ICML*, 2006.

[13] Wang, C., D. Blei, D. Heckerman. "Continuous time dynamic topic models." *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2008.

[14] Wang, X., A. McCallum. "Topics over time: A non-markov continuous-time model of topical trends." *Conference on Knowledge Discovery and Data Mining (KDD)*, 2006.

[15] Wei, X., J. Sun, X. Wang. "Dynamic mixture models for multiple time series." *IJCAI*, 2007.

[16] Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, L. K. Saul. "An introduction to variational methods for graphical models." *Learning in Graphical Models*, 1999.

[17] Corless, R. M., G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth. "On the lambert w function." *Advances in Computational Mathematics*, 5:329–359, 1996.

[18] Radev, D. R., M. T. Joseph, B. Gibson, P. Muthukrishnan. "A Bibliometric and Network Analysis of the field of Computational Linguistics." *Journal of the American Society for Information Science and Technology*, 2009.

[19] Joseph, M. T., D. R. Radev. "Citation analysis, centrality, and the acl anthology." Tech. Rep. CSE-TR-535-07, University of Michigan. Department of Electrical Engineering and Computer Science, 2007.

[20] Schmid, H. "Probabilistic part-of-speech tagging using decision trees." In "Proceedings of International Conference on New Methods in Language Processing," 1994.