The Preparation Of The Topic Model

Rachel Sagner Buurma
rbuurma1@swarthmore.edu
Swarthmore College, United States of America

Replacing "How something is made, with a view to finding out what it is" with "How something is made, with a view to making it again" – the Essence with the Preparation – is linked to an option that's completely antiscientific: in reality, the starting point of the Fantasy [of the critic's writing of a novel] isn't the Novel (in general, as a genre), but one or two novels out of thousands.

-Roland Barthes, *The Preparation of the Novel,* Session of December 9, 1978, 13.

The Literariness of Topic Modeling

This short paper reports on the progress of my attempt to construct a reading of topic modeling using state-of-the-art literary criticism. I argue that dominant digital humanities understandings of topic models assume some of the characteristics of literature most essential to twentieth-century criticism – counter-factuality, a mediated form that is ultimately separable from aesthetic characteristics, and an efficient, self-enclosed, total form. More specifically, I show that topic models also tend to be read by digital humanists according to the assumptions, protocols, and caveats we accord to the interpretation of realist fiction. The result is that while often revealing and productive, many digital humanists' uses of topic modeling are indebted to assumptions about the literariness and fictionality of topic models that we have yet to fully understand. Drawing on work by Stephen Ramsay, Johanna Drucker, Alan Liu, and others that theorizes continuities between the values of literary criticism and computational processes, I suggest that we temporarily set aside the idea that topic modeling reveals the "contents" of a set of novels (or of any other corpus). Instead, drawing on Roland Barthes' late work on *The Preparation of the Novel*, we might rethink topics as preparatory notes written by no one, as an imaginary archive whose contents furnish a productively alienating, too-perfect map of the novel's preparation. In *Preparation,* Barthes moved away from his earlier work's emphasis on totalizing interpretations of literature's meaning to think about models of the text that allow for a more partial and slow view of the process of meaning creation (See Buurma and Hefferman, 2014) Topic modeling has the potential for helping us towards a Barthesian reimagination of the novel's reader as the novel's writer, of the search for the fantasy origins of a novel as a method that pulls us away from formal totality and a form-content divide. While this reorientation comes out of literary studies, I also suggest that it might have applications for more instrumental uses of topic modeling outside the realm of the humanities, in which assumptions about topics as equivalent to a document set's "contents" also tend to draw on our conventions for reading realist genres.

Fictionality and the Topic Model

The past few years have seen the rapid popularization of topic modeling among humanist scholars in general, and among scholars of literature in particular (see Blei, 2014; Erlin, 2014; Goldstone and Underwood, 2014; Laudun and Goodwin, 2013; Meeks, 2013; Jockers and Mimno, 2013; Rhody, 2013 and 2016; and Tangherlini and Leonard, 2013). The literature on topic modeling abounds in stern and salutary warnings about the limits and dangers of topic modeling for humanistic study. One can read about the dangers of introducing algorithmic black boxes into literary research, the concern that literary scholars are unprepared to fully (or even partially) interpret the topic models and their related data, and the worry that they fail to understand even the interpretive choices made during corpus preparation. Part of the worry derives from a larger assumption that topic modeling "reveals" the "contents" of novels. We assume that literary critics dipping their toes into topic modeling will shed their traditional interpretive caution in the face of the algorithm's authority, and will misunderstand the un-semantic nature of topics or accept meaningless correlations as meaningful. I want to suggest that all such warnings are relevant only given a very limited understanding of what a topic model is, its imagined relation to the corpus from which it derives, and the goals of the model's interpreter. These warnings do usefully help us think about some of the seemingly inconsequential interpretive choices we make when we choose chunk and clean documents, apply stoplists, select a number of topics to train, and, most importantly, assign semantic labels to unsemantically

generated topics. And yet these warnings assume either that topic models aspire to be mimetic maps of the corpuses they model or that technologically unsophisticated interpreters of topic models imagine that this is the case. Schmidt (2013) warns, for example, that "simplifying topic models for humanists who will not (and should not) study the underlying algorithms creates an enormous potential for groundless — or even misleading — "insights."" He worries that a pair of assumptions about topic models – that they are "coherent" and "stable" – "let humanists assume that the co-occurrence patterns described by topics are meaningful; topics are useful because they describe things that resemble "concepts," "discourses," or "fields."" He is worried, that is, that the appearance of semantic meaning we find in "good" topics will seduce humanists into thinking that they have discovered the "contents" of novels – whereas what topic modeling really offers us is exactly a non-semantic machine indexing of a set of texts about which our approaches tend to be based on ground assumptions about semantic meaning. This is not surprising; the assumption that topic models are a realist genre is pervasive in literature on topic modeling, literary and otherwise (for example, Airoldi et al, 2015, describe good topics with the example of "trout fish fly fishing water angler stream rod flies salmon…" explaining that the topic "is specific. There is a clear focus on words related to the sport of trout fishing. It is coherent. All of the words are likely to appear near one another in a document. Some words–water, fly– are ambiguous and may occur in other contexts, but they are appropriate for this context. It is concrete). Yet if we relieve ourselves of this constraint and instead substitute a more plausible frame – the topic model's fictionality – we will be able to enjoy a wider range of relations between model and corpus.

In place of assuming that topic models belong to the realm of realism, then, we might pay more attention to the generative uncertainty of topic modeling and to its literal fictionality. Topics are probabilistically-created formations, and the algorithm that generates topic models is based on the enabling--but crucially, counterfactual--"assumption that documents have multiple topics." (Boyd-Graber et al., 2015). By looking at the documents we offer it, the algorithm generates topics that, in given proportions, compose each document. (Or, rather, it generates the probability that a certain percentage of words in every given document were generated by a given particular topic.) Topics, of course, don't actually exist prior to the documents that generate them; they don't actually exist independently

in the same way the documents at all. They are, in a certain sense, fictions. Topics are things that might have existed – but didn't! - given the existence of the document set in question. While we can and sometimes do relegate this fact to the realm of methodology, the fictionality of topics is crucial to remember for any literary-critical uses of topic modeling, for it reminds us that these models offer us a view of our document set radically at odds with any other more literal sources of a novel we might use – such as an author's notes towards a novel, or a catalog of the virtual or actual library of books a novelist brings to the writing table, or even the looser sense of social "discourses" that exist prior to novels and which we might imagine in part "composing" a novel. As Boyd-Graber et alia note, "Topic models are based on a generative model that clearly does not match the way humans write. However topic models are often able to learn meaningful and sensible models." (2014: 15). Using a few targeted examples drawn from topic models of corpuses of nineteenth-century novels of varying sizes and comparing them to some examples of nineteenth-century novelists' notebooks, I suggest that reimagining topic models as fictional notes might be not just a theoretical exercise but a practical way of conceptualizing the relation between topic model and corpus.

Bibliography

Airoldi E M, Blei D, Erosheva E A, Fienberg S E (eds), (2015) *The Handbook of Mixed Membership Models and Their Applications*. Boca Raton, FL: CRC Press, Taylor & Francis Group)

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM* 55, no. 4 (April 1, 2012): 77. doi:10.1145/2133806.2133826.

Blei, D. M. (2014). Topic Modeling and Digital Humanities. *Journal of Digital Humanities* (April 8, 2013). http://journalofdigitalhumanities.org/2-1/topicmodeling-and-digital-humanities-by-david-m-blei/

Belvins, C. (2010) Topic Modeling Martha Ballard's Diary. April 1, 2010. http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/

Boyd-Graber, J., Mimno, D. and Newman, D. (2014). Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements. In: Airoldi E M, Blei D, Erosheva E A, Fienberg S E (eds), *The Handbook of Mixed Membership Models and Their Applications*. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2015.

Buurma, R.S. and Heffernan, L. (2014). "Notation After 'The Reality Effect': Remaking Reference with Roland Barthes and Sheila Heti." *Representations* 125:1 80–102. doi:10.1525/rep.2014.125.1.80.

Chang, J, Boyd-Graber, J.L., Gerrish, S. Wang, C, and Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. In *Advances in Neural Information Processing Systems* 22, 288–96.

Erlin M (2014). The Location of Literary History: Topic Modeling, Network Analysis, and the German Novel, 1731-1864, p 55-90. In Matt Erlin (ed. and introd.) and Lynne Tatlock (ed. and introd.), *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*, Rochester, NY: Camden House.

Goldstone, A. and Underwood, T (2014). The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us. *New Literary History: a journal of theory and interpretation* 45:3.

Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.

Jockers, M.L. and Mimno, D (2013). Significant Themes in 19th-Century Literature. *Poetics* 41:6, 750–69. doi:10.1016/j.poetic.2013.08.005

Laudun, J. and Goodwin, J. (2013). Computing Folklore Studies: Mapping over a Century of Scholarly Production through Topics. *Journal of American Folklore* 126:502, 455-475.

Meeks, E (2013). The Digital Humanities Contribution to Topic Modeling. *Journal of Digital Humanities*, 2:1. http://journalofdigitalhumanities.org/2-1/dh-contribution-to-topic-modeling/

Ramsay, S (2011). *Reading Machines: Towards an Algorithmic Criticism.* Urbana, Chicago, and Springfield: University of Illinois Press.

Rhody, L.M. (2013). Topic Modeling and Figurative Language. *Journal of Digital Humanities*. http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/

Rhody, L.M. (2016). The Story of Stopwords: Topic Modeling an Ekphrastic Tradition. *Digital Humanities 2014,* Lausanne, Switzerland. Accessed January 3, 2016. http://www.lisarhody.com/the-story-of-stopwords/

Schmidt, B. M. (2012). Words Alone: Dismantling Topic Models in the Humanities. Journal of Digital Humanities, 2(1), http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/

Tangherlini, T.R. and Leonard, P (2013). Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research. *Poetics* 41:6 (December 2013): 725-749. doi:10.1016/j.poetic.2013.08.002.