Mikko Hyyryläinen

Type: *Writing an response to data science article of your choice discussing its methodological choices: 5 points per each response*

Link to the article:
https://www.sciencedirect.com/science/article/abs/pii/S0304422X19300154

This is a response to Luca Pareschi and Maria Lusiani's (2020) article *What editors talk about when they talk about editors? A public discourse analysis of market and aesthetic logics* which is a pre-print phase in *Poetics* journal. The research is discourse analysis that uses the methods of topic modeling and multiple correspondence analysis. As a side note, like the response paper to the other article, this is about topic modeling since it is the machine learning method that I am most interested about. I also find it interesting that an article that identifies itself with discourse analysis uses machine learning (and statistical) methods since my impression about discourse analysis is that it is interpretative analysis of texts where different nuances are important and are hard to find with computational methods.

The article starts from the observation that "there has been an institutional change going on in cultural fields, one in which the market logic gained more prominence and started interplaying with aesthetic logic in many ways" (1). The tasks that the authors do are: 1) find the "manifestations of market and aesthetic logics" from "public reconstructions of their work in a cultural field" and 2) find out "how are these discursive reconstructions related to field characteristics?" (2). The domain they analyzed was Italian literary field and the data was "87 interviews with [literary] editors discussing their work" (2). The theoretical background of the article is in Bourdieu's conception of fields and cultural production and also the idea of discourse as something that constitutes institutions (or fields and their logics).

How the authors frame the data they use, that consists of interviews in Italian magazines, is that they "consider interviews as narrations actively constructed by those interviewed" that are cases of "editors' public reconstruction of their work" (4). If we accept the fundamental principles of discourse analysis, this framing makes sense and considers the context where the data is created in an adequate manner. What I find interesting is how the authors describe what topic analysis is about. Their description could be considered "non-technical" where the interpretation of the topics is considered to be part of the topic modeling method (which it is in social scientific practice, but if we consider topic modeling as a machine learning method, it does not necessarily include the interpretation part) and the way that the topic modeling is described is quite superficial. The authors also used "state-of-the-art software Mallet" for the topic modeling and although they discussed removing the stopwords, they did not discuss stemming or removing the most frequent (non stop-)words (which seems to be customary). This made me wonder if Mallet did these things by default or did the authors consider these phases at all.

The process of how they ended up with a topic model with six topics is quite brief (this may also be because of how an article is structured in this type of social scientific journals where the results and conclusions based on it are more interesting than the description of the process of finding the right model or mathematical description of the method). The method of analysis that authors used seemed to be that they used in-depth readings of the texts, topic modeling and also re-reading the texts in relation to the topic modeling results. This makes me ask why they used topic modeling since the number of interviews was 87 and it could perhaps be said that this is a bit too small set of texts for doing topic analysis in a meaningful way.

The second part was about using "MCA to relate these meanings [that editors produced in the interview]" to the "structuring features of the field" (9). These structuring features of the field were 1) size of the publisher where the editor worked, 2) geographical location, 3) gender of the editor and 4) source of the interview (9). It seems that the authors considered that these features influenced the way the editors structured their work in the interviews. They found out, for example, that Topic 5 (intellectual status discourse – or topic) is "close" to medium publishers and Rome (10). Hence, it was an interesting combination of two non-qualitative methods that provided an answer about how discourses related to different attributes of the field.

Although the article was interesting and it integrated methods and theoretical ideas, I think that there were some aspects of it that were not that strong. The biggest thing that I wondered was why they decided to use topic modeling. As the authors themselves say, their data consisted of a "manageable number of texts" and I did not understand why they chose to use topic modeling to analyse the data. Especially since the way they described the process of topic modeling created the impression that they were not that familiar with the method and its background. Even if it would be considered somehow supplementary to the in-depth analysis, its role was too large in the article. One problem was also that although the article was about something that I would define a process, "changing relationships between market and aesthetic logics", it only took one snapshot of the field in question, and at least I would consider that for researching "changing relationships" it would be necessarily to take at least two snapshots and compare them.