Mattia

Tue, 9/13 4:47PM • 21:07

**SUMMARY KEYWORDS**

validity, analysis, validation, methods, measure, paper, validate, text, coding, supervised, provision, construct validity, standards, eu, research, complexity, parsing, delegation, manual, governor

00:00

Okay, then I will start recording now. Thanks a lot. So yeah, as I just said, Could you then maybe start with just describing how you think, like, what do you think that measurement validity is and what it means to you and your work?

00:21

Yeah. So I mean, relatively speaking, like a measurement validity is like you need to, you know, you measure that you actually, you see what you are major, so you get a fair measure of ordering a concept. So you operationalize it, you have a concept, we operationalize it, then you measure it, in this case with text analysis, as an example. So you want to make sure that you measure what you have in mind at a conceptual level. So a classical example would be, you know, like you have this concept of populism and how you measure it, do you just count the number with the frequency of tokens that come up in the sense of like, the people late corrupt, and so on, do you look at the rhetoric, so whether they emphasize some things, whether they talk more easily, and so on. So there are different way to measure a concept. And both are legitimate, but we want to make sure that there is a validation. So, validation, like what I use, it protects analysis is literally, you know, my idea of text analysis is that, like, it should be used only in those occasions where you cannot get information otherwise. So sometimes you can get the information you need, from, you know, like, submit the data, so who's speaking, who's the speaker, who's the signatory of a deal, you know, like, and something like that. So what we want to make sure is that we use text analysis only when it's needed to get an extra layer of information. That extra layer of information anyway, needs, we need to make sure that we measure what we have in mind. So in that case, that's where validation is important. It is, to me particularly difficult index analysis, especially for this feature, because, you know, since we use text analysis, to get extra information from other observational data we might have, it becomes very difficult to have ways to check the validity. So one of the more common most common ways and also one we use is face or construct validity. So in our political analysis paper, we say, Okay, we have this new method to measure the negation of power to the government. So in the US state, we have the corpus of all the US states session lows. And what we want to see from there, we want to have a measure of delegation. What how we do it, we count the number of provision that delegate powers to the governor. We did that and then we divided by the number of provision that constrain the governor. So the governor is allowed to do that delegates power, the governor is prohibited to do that, that restraint. Okay. So there, we have a measure of delegation. So how do we, how do we validate it? First of all, is face and constant validity, as I said, so if that is the case, if our measure makes sense, we might see some things. So the first thing would be that we might see more than negation when the two chambers and the governor come from the same party. Right. And that's what we saw. That's more next boss delegation. That's construct validity. Or you want to validate it with time, the golden standard would be manual coding. So what we do in this case is we take Fabula, Frankie no dataset, which is more, which is focused on the EU. We run our code on the EU, and we see that you know, like there is high correlation in terms of those directives that delegate power to the commission, according to manual validation, manual coding and according to our automatic coding, so the optimal way to do validation when it comes to when it comes to effects analysis will be more an exam, the one and you know, the gold standard can be manual validation.

05:09

And also an exposed one, which is like face validity of construct validity. So what, in an inferential statistics setting, what we would see, were our measure, you know, like make sense? And those are the two ways that I have used in my research.

05:32

Yeah, did you also get the feeling that this view is somehow accepted within the research field, also, maybe in connection to different kinds of methods? And in I mean, I've read your paper this morning about this syntactic parsing, which you just talked about? Do you also, like have any ideas on the validation of different types of methods or, for example, looks supervised or unsupervised methods in this regard?

05:59

No, like the equivalent favor like JSON, and as the populace and Tony Bertelli, they have a paper where they do practically what we do, but with machine learning. So in that case, what I mean, they have a subset of of provision that they know according to manual coding that delegate power in that case, I think, to the commission, because it's the use setting. So they train the algorithm on that, and then you know, like, they do their analysis, I think it's always good to keep our meter, our method is not supervised. So we have on volunteers, then we match them with extraction rules in our tests, and then we just count the number of provision. It would be nice to also Yeah, it would be nice to validate our method with more supervised one, especially because nowadays there are the syntactic parsing that we're using spacey also allows to sorry, give me a moment also allows you to have more custom semantics, as in like, you, you can create your semantics like we do in terms of in terms of the words you would say for the app, the note the main verb, the subject and so on. And then you can actually, you know, like, you can train your algorithm on that. So, because the syntactic parsing we use, in the end, it is supervised. So, it is trained on those three banks, they tell you, you know, like, if a noun is is followed by a verb, you have certain probability that that is the subject. Okay. But that is based on natural language as in how we use the language as people. But that can be done in a sort of more institutional language. So, you might already have your custom semantics they tell you, Okay, every time the word governor is followed by allowed prohibited or a modal, that is a provision that just isn't that. So I Yeah, both in terms of tax and tax, both in terms of the tax analysis. And in terms of validity, I think that the one one possible avenue for the future is mixing more supervised and unsupervised techniques.

08:59

Great. I mean, before coming to this other part, it has one last question in terms of religion validation. So, one thing we've observed and also talked a lot to people about is within the research process, that there's like very high degree of research of freedom of research, degree degree of freedom for researchers. So like a lot of lot of steps to talk about, or to think about, do you have any experience on validation practices, which you do within your research process from gathering data, descriptive statistics and then analyzing data or your method or validating your method, which is not really talked about that much in papers, because in papers, your own report, like the end results, the post hoc validation of your method?

09:50

Yes. So I mean, first of all, I have to say that a good portion of the paper that validation file that you see or even don't see, sometimes it comes from the peer review process. Sometimes you have reviewers asking for validation or more validation. Because as again, there is no, especially with text analysis, there is no standard template on how to do it. So that was honestly the case for our paper, we thought the Construct validity was enough. And then rightly, a reviewer said, okay, but can we provide something more exacting. And so we've got, we asked funding from, you know, for the data set, and so on. So that is also a community and they. So we got a really good reviewer who really understood the methods, the, the potentials and the limitations. And so we manage, and actually, the paper now, will, is much better than it used to be without that piece of a lesson. So you also need that sort of evader just because we didn't think about it makes totally sense, but we didn't think about it. So compare is one way that validation can be nowadays, even I do it when I receive, you know, like a paper that does text analysis. I always check that app and I tried to ask for more replication or materials and so on. So we're currently working with a colleague of mine on a paper measuring complexity in EU directives. And they're what you're not going to see in the final paper. But what we have done is we read some EU law textbook on how to write EU legislation. So those are meant for legal drafters, people that actually write legislation. And those are written by lawyers, I mean, legal scholars, but also backed by people who works in the EU. So in the EU, there is an old DG that deals with that. And so we read that. And then we knew that some directives, like the parental leave directive, is known for be complex, and is known for creating issues for the implementation. So that was also against the whole art of face validity and constant validity that we needed to have in mind, because I did my research. When other people were starting to work on this complexity idea, especially measure with text analysis and NLP techniques. But we really want to know where the complexity comes from. So we need to be to be clear about the concept of complexity. So in that case, I mean, we define complexity as being like, you know, like creating some issues for those that need to implement legislation. So a more complex EU directive is less likely to be implemented properly at member state level. Because he leads because it can be interpreted in different ways strategically unknown, he can, you know, like, leave way to exemption exception and so on. And, but that he came, you know, like that, that came actually by studies, those EU law textbook on how you write legislation. So and that is all a part, both in defining the concept of validity. And second in measuring it that you don't see in the paper. And the other part for your question about like validity measurement is exactly that. So we are gonna have some former validity measures in the paper, but all these far, you want see in the paper actually going through those EU directives, especially the most complex one, and have an idea of what we're looking at.

14:44

Okay. And then, I mean, you already talked about, like, you mentioned reputation materials. For example. Do you have any further suggestions or ideas on how this field of research can advance or validation practices? In general, for for text methods

15:04

in terms of replication, no, yeah,

15:06

no, no reputation was just one thing. I mean, I think you mentioned reputation, like the publication of replication materials. And I think you also mentioned standards, or like, you said that there is no general guidance on how to validate methods based on texts.

15:21

So I mean, like, it would be nice to I mean, it doesn't have to be for like, proper protocol, or you know, like, a template, it would be nice to have an idea in the community of what is considered a legitimate way to validate your methods when it comes to text analysis. Because for reasons for these constitute validity, being exposed, the boundary between validation and the analysis part are really blur. So in that case, you know, like, we might be good to have an idea of a sort of, you know, like, a year or key of validation approaches. So validating with manual coding, survey, expert survey, or you can also do those Leastwise comparison. So what we are going to do with the complexity one since is is, is a concept that it is difficult to manually validated, what we're gonna do, we're gonna have a list of provision, and we're gonna ask no expert to say, Okay, which one of the two is more complex, according to you. So it would be nice to have a year or key of approaches that is accepted in the, in the community so that when I cannot get the first one, let's say, which won't be like, you know, like biannual coding by asker, I can go to the second one, which is least wise comparison. And then I can go to the third one, which can be construct validity, when I really cannot do anything else, because to validate is fine.

17:26

Okay, yeah. So then what do you think prevents the that's, that's also the last question, what prevents the establishment of such standards? Or frameworks? Do you think that there's like any reason for that, for example, the variety of research context of it of methods like text data in general, because in other disciplines, there are, for example, if you look at survey research, there's like a very, like, somehow accepted validity or validation steps when you, for example, construct a scale.

18:03

Yeah, I think that we text analysis. So a bit of that is, I mean, like, strategic capacity, that's for sure. The other part is the nature of text analysis. And also, because you know, like the end of the day, feel have some text and you download the say, container, it takes nothing to come up with some topics, right, or to come up with frequency distribution of tokens, or even co occurrence, or something like that. So I think it is a base between the fact that text analysis gives an extra layer of information, and that is the advantage of dictionaries, and also the accessibility of tools without the necessary methodological background. So isn't mixed up. And then obviously, you have some strategic capacity in terms of the find that you really don't want to look into what you're doing the black box, otherwise, you're gonna find something you don't like?

19:21

Yeah, great. Is there anything else you would like to add to this whole discussion, Texas data validation? Anything?

19:31

Again, I really think it is really important, it would be nice to start. It would be nice to start with the stimuli to start at really early in the community at PhD level. So if you have a PhD student that works on text analysis, saying with statistics, the first step will always be you know, like to read play Okay, I saw, you know, like some main papers in the areas making sure I mean not to double check other people were but also because you are aware of how things are done. And you became aware of the best practices earlier. Gauze, you know, like nowadays is a pretty established set of tools and methods. And nowadays we more or less have an idea of how to validate things. But I think it's, again, a community endeavor that needs to come from the bottom up and even people like me by especially my quarter, you know, fine, we have our ways of doing things, but it would be nice if like, you know, we start with the training PhD students to get even more transparency, more replicability and have a set of standards for them.

21:03

All right, yeah. Then I will stop the recording.