

Text analysis interest group inaugural meeting.

JSM, Denver, CO, 30 July 2019

20+ attendees

Stas Kolenikov presiding; Tommy Jones taking notes.

Motivation in creating the group

Introducing on the origins of the interest group and ASA process:

- petition required to create an interest group, outlining purposes and potential membership;
- petition was signed by ~140 people, almost 5X over the 25 required.

What is an interest group? Similar to a section, it brings together people of similar interest. Unlike sections, it does not have as strong of an "official" representation with ASA. Distinctions:

- sections vote in the Council of Sections; interest groups are non-voting members
- sections have specific allocations of invited sessions at the JSM (1-4) and topic-contributed sessions (up to ~10); an IG has one topic contributed session, and can submit proposals for an invited session
- both sections and IGs can co-sponsor sessions they like, after the program is formed by the primary sponsors of sessions/presentations
- section members must be members of the ASA; IG members don't have to be ASA members

One possible venue is to eventually turn an IG into a session so that we could then sponsor invited sessions. (Interest groups cannot do this.) What IGs can do, is co-sponsor sessions. This makes it easier to search for sessions with text analytics-related material. We can also do a topic-contributed session. We have to propose this to ASA.

Stas hopes that this interest group will span statisticians, computer scientists, social scientists and more.

Group formalities

For ASA purposes, activities of the group are governed by the Charter (which is mostly a boring document).

One important part is the set of officers. We need to find officers. Officers interact with the ASA to facilitate two-way communication between ASA and the members.

Officer slots

- Group chair - responsible for most activities of the group, sans those outlined below for other officers
- Program chair - responsible for coordination of submissions to ASA meetings; identify sessions at JSM that are of interest to TAIG members.
- Secretary/treasurer - take notes, membership roster, manage budget

The group budget is \$1,000/year from ASA to use. You have to apply to use it. It's not automatically allocated.

Stat's guesstimate: chair is 2 - 3 hours per month responsibility. Secretary/treasurer is responsible to taking minutes and tracking other communication. And as treasurer they have to track and manage money. These are likely to come in bursts at random times. The Program Chair needs to put down some work in the fall when the invited and topic-contributed program is created to identify speakers; and in late spring to identify additional sessions of interest and request the ASA meetings staff to get them added as co-sponsored by TAIG.

Question from the audience: Clarification on purview of text analysis interest group. Does this include all aspects of NLP? Social media analysis? Tagging images? Stas' expectation is that the group will do all of that; as this is the first organization of this kind at the ASA, there is little point to be restrictive at this moment.

- Idea at this point is to have the group set the boundaries.

Stas will collect names of people interested in officer positions.

Stas was nominated by David Banks to be the chair. James Keiser seconded. There was a voice vote. (Not fully clear how binding that is. See notes below.) Stas expressed that he is only intending to fill this role for a year or so.

Tommy Jones suggested that for officer positions, we do that offline and via email etc., so that others who could not join in the room would have an opportunity to put their names forward.

Stas will look into the way to set up a proper election ASAP.

David Banks suggested that instead of setting up a student award, we do something more "daring" for use of that money. Examples include: a challenge (analysis/coding).

Short-term plans

- JSM 2020 participation: we discussed ideas for what a topic-contributed session could be.
- Stas reported that there is substantial interest from the ASA on the topic of social media, so we could offer a short course on social media analytics.
- The current ASA president has an initiative to solicit statisticians' input on disinformation. "Text analysis must have a role in that." Something along those lines may make interesting session fodder. But also it can bring us to the table at higher level discussions at ASA.

Question: is there a spot in the charter for a *student rep*? Adding one would be a reasonable thing to do though the charter isn't currently set up for it. The charter is a living document.

Davide Marchette proposed to put together a group to look at the charter and get one set that the group can agree on.

Another potential officer position to consider: a publications officer to help disseminate information.

One way to approach this: Have the elected officers nominate additional ones (e.g. student rep) to lower the burden on members as well as have them be "trial" positions. i.e. does this make sense.

Suggestion: We could offer short courses and that

Opinion: statisticians in this field often do the work that CS people are doing, not focused on uncertainty quantification. How do we calculate power in a textual context?

Future section potential?

Do we want to create a section? If we are going strong: there's a good argument for it. However, if we do create a section, we may cut off other disciplines.

Arguments for and against. However, we have to wait three years to propose. So, we can watch and evaluate and then revisit this question in 3 years' time.

Introductions with people in the room

(Following a random walk)

Jason Brinkley - works with Stas at Abt. Learning NLP on the job and getting interested.

Richard Penny (Statistics New Zealand) - argued there isn't enough statistics in text analytics and unless ASA pushes it, it'll be all computer scientists

Lisa Frehill (Institute for Museum and Library Sciences) - wants to build bridges to other disciplines that have done this type of thing for a while.

Jian (?) Sun - Case western university but about to move to George Mason university - interested in electronic health records

Qiuyi Wu - Phd Student at U Rochester - Master's thesis used text mining as related to music

Todd Sanger (Eli Lilly) - runs group at EL that does text analytics. Has stats PhD, but not many statisticians involved

Shane Bookhuttz - Phd student at VA tech - interdisciplinary research on polarizing information

Scott Comulada - Faculty at UCLA - hasn't done text mining but is very excited at creation of group. SMS text involved in his research and could really use methods to analyze it

Anne Parker - business analytics lab at IRS. Text analytics is new to them but they have a vast trove of textual data that needs analyzing

Kelly Townsend - Statistician at John Hopkins Applied Physics Lab doing text mining on maintenance reports on the Navy

Karl Pazdernik - at Pacific Northwest national lab - does lots of text analytics and can't talk about what they're doing exactly

Jordan Rodu - UVA - frustrated with lack of statisticians in the area and lack of uncertainty quantification in the field. One thing this group could do: do a competition with an industry data set on UQ

Xu Yang - SAS - developing a platform for customers to do text mining. Very interested in text mining across languages

Richard Schwinn - Small Business Administration; they have an internal(?) text mining package. Hoping to copy and steal all of our ideas.

Joseph Willard - PhD student at U Michigan doing text research

Stas sent attendance sheets to collect the names and email addresses