

Coding Methodology for “Much Ado About Gender: Current Practices and Future Recommendations for Appropriate Gender-Aware Information Access”

Author: Christine Pinney

Coding Papers Method:

- copied title, link, and venue into spreadsheet, then determined what the paper type was.
 - ACM provided labels that were crosschecked with the accepted papers on each individual conference website to ensure that results were consistent.
- first I used ‘cmd-f’ to search for the keywords, which would allow me to see how often the document mentioned each keyword
 - provided initial glimpse on what to expect as far as gender variable usage, helped filter ‘yes’ from ‘applied’ and/or ‘no’
 - location was also important (some only had hits in citations, introductions, tables, related work...)
 - if this revealed possible gender variable usage, I searched for other terms (female, woman, etc.) to check for other discussion of the idea
 - Determined if gender was being used in terms of describing the dataset, or if there was more relevant discussion.
 - looked for some kind of inference or conclusion made about the gender variable
 - If I found that there was discussion that warranted further reading, I then read the paper more thoroughly to determine how to properly code it.
- If no:
 - provided quote of one instance with keyword(s) to give context
 - gender distribution:
 - looked for percentages, () with numbers
 - demographic data:
 - collected via questionnaire, survey, etc.
 - contextualize argument:
 - if gender was used as an example variable or other studies with gender variable are referenced, usually intro or related work
 - tag:
 - keyword shows up as a tag in a larger query or label of an image
 - future work / footnote / citation
 - attribute / user profile:

- if mentioned (often with age, education, etc.) as some kind of user attribute or as a part of a larger user profile (but gender is not discussed further)
 - Different from vectorized/grouped where the vector/group is used within a model (this is applied)
- If applied:
 - The difference between 'no' and 'applied' is the implication— implicit or explicit— that gender is a variable that the study's model could be applied to, but was not.
 - or that gender is a variable of part of a larger vector and/or group that is used, but gender is not discussed individually.
 - vectorized/grouped:
 - Gender is a variable that is used within a larger group and/or vector but is not discussed individually
 - The difference between vectorized/grouped and user attribute/profile is pretty much just keywords and or language used, but the goal is pretty much equivalent
- If yes:
 - referent: provider, users, subject, annotators
 - is the gender variable applied to the user, or is it applied by annotators to some subject (images, queries, etc.)
 - multiple referents
 - Is the gender variable being applied to more than one referent?
 - gender categories: usually based off of what was explicitly stated in the paper
 - non-binary: 'ack' if the author(s) make some clarification about how gender isn't necessarily binary, but they are using binary gender because ultimately it's easiest or what is provided by the dataset.
 - gender determination:
 - analyzed dataset/methodology paragraph for determination (e.g., "questionnaire", "survey")
 - goal:
 - most often identifiable within the intro/conclusion
 - contextualize results includes user experience surveys
 - bias/fairness: was there any discussion about how the findings are imbalanced in terms of gender? or suggestions/recommendations towards making the system more equitable?
 - most often identifiable within the intro/conclusion