Assignment 2 Write Up

Eric Brigham   Hunter Mundy   Sophie Garrett

The data used for this assignment was gathered from a *BioScience* paper titled "Internet Blogs, Polar Bears, and Climate-Change Denial by Proxy". The paper examined how anthropogenic global warming (AGW. Global warming caused by human activity) in the regards to arctic ice levels and polar bears were covered by scientific blogs and papers versus climate change sceptic blogs. The paper's data consisted of 90 blogs total, with 45 scientific blogs and 45 denier blogs, along with 92 papers.

We decided to examine the paper's data in order to find common co-occurrences of specific positions. We began by trimming the data in Excel. The original data contained several extraneous columns called symbol, color, and code. These were used in the paper's graphs and for simple categorization respectively. After we removed them, 9 columns remained:

1. yes/no Crockfrod (whether the source mentioned Crockford, a supposed "scientist" and known denier)

2. yes/no ice decline (whether Artic Ice levels are rapidly decreasing and are caused by human activity)

3. yes/no ice stable (ice levels are not decreasing rapidly or are even increasing)

4. yes/no ice natural (the reduction of the ice, if there is any, is due to natural phenomena and is impossible for humans to predict or affect)

5. yes/no bear threatened (polar bears are threatened due to AGW and polar bears might go extinct if action is not undertaken)

6. yes/no bears adapt (bears will adapt, regardless of what happens)

7. are bears okay yes/no (bears are not threatened from any form of AGW)

8. position ("pro" if the source believes in AGW, "contra" if it doesn't)

9. source (blog or paper).

Note for test results: I'm aware the column values are worded weirdly. This is how they showed up in the data and I decided to use the same wording to save space. Reminder "pro" refers to the scientific side and "contra" refers to the sceptic side.

During testing, it was discovered that a minsupport of 148 and any confidence level gave us no results. Any minsupport from 135 to 147 merely returned one frequent itemset, "no_crockford", thus revealing the absence of mentioning Crockford to be the most common column value. Additionally, any support value lower than 120 tended to produce so much data that it seems unreadable. In the support range of 120 to 134 and any confidence level, the most common columns were always "pro" aligned ones. Although this could be due to the abundance of pro-science sources (discussed below), the lack of any "contra" columns indicates that

"contra" positions rarely talk about a range of topics similar to how scientific positions do. The results suggest that "pro" papers outright state ice is declining and that polar bears are in danger and then go on to explain that views contrary to these are false. This explains why "pro" columns appear so often together. In contrast, the results imply that "contra" sources either pick one view and argue for it or, as noted in the paper itself, merely attack scientific sources.

Using a support value range of 129 to 133 and confidence of 0.4 revealed the most common items scientific column values. They were, in the order of least common to most common: "pro", "yes bear threatened", "yes ice decline", and "no crockford". The columns "yes bear threatened" and "yes ice decline" appeared most often together, while "pro" and "no crockford" appeared in almost every frequent set.

Although the position of "no bear ok" is a scientific one, it rarely shows up with other scientific positions. This is interesting as "no bear ok" is essentially the same as "yes bear threatened", suggesting "pro" blogs and papers state in plain terms the precarious situation the polar bear is in. Alternatively, this shows that the paper's authors didn't need what basically amounts to two columns saying the same thing.

Relatedly, there was no frequent mention of "no bear adapt". This suggests that not only is there not many blogs/papers that mentioned the topic, "pro" blogs/papers didn't consider the idea important enough to even say it's not true. Another infrequent column mention is of "yes/no ice stable" and of "yes/no ice natural", reconfirming the fact that "pro" sources focus on blatantly saying that the ice is melting due to human activity. "Pro" sources don't waste time discussing the supposed "trues" put forth by "contra" sources, believing the overwhelming amount of data in support of AGW invalidates any other possibility.

Using a support of 134 and a confidence of 0.2 revealed that the most common columns were "no crockford" and "yes ice decline". As this is the highest support level before getting nothing but "no crockford" for the most frequent item, this shows that almost every scientific source agrees that AGW is real.

After reviewing the data, I recommend that the authors of the paper either reduce the number of scientific papers used, increase the number of "contra" papers used, or remove the papers altogether. Only 6 out of the 92 papers are labeled "contra", biasing the results to favor the scientific community. Since the paper already gathered an equal number of scientific and sceptic blogs, its odd the authors didn't do this for the papers as well.

The authors could also do away with the "no/yes bear ok" column since it is functionally the same as the "yes/no bear threatened" column. Saying that polar bears are threatened implies they're not okay and vice versa, thus making the column irrelevant.