Methods of Computing Committee Balance

Balance is a number describing the committee of 10 politicians that a user chose. There are several ways to compute this value, described below. See Balance.csv to see the data for each participant.

AD: AD (stands for "attribute distribution") is a way of quantifying committee balance using the Attribute Distribution bias metric for each attribute of the data. It ranges from 0 to 1, where 0 represents a perfectly balanced committee, and 1 represents a highly imbalanced committee. The underlying computation for this can be seen in Wall et al.'s paper on the bias metrics[1].

[1] E. Wall, L. Blaha, L. Franklin, and A. Endert, "Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics", IEEE Visual Analytics Science and Technology (VAST), 2017.

Diversity: Diversity quantifies balance with respect to how diverse the committee is for each attribute of the data. It ranges from 0 to 1, where 0 represents a perfectly balanced committee, and 1 represents a highly imbalanced committee. According to this viewpoint, the ideal committee has an equal number of people across each possible value. For example, a gender-balanced committee would have 5 men and 5 women. An education-balanced committee would have 1-2 people of each degree level (7 possible values).

Let $x_1, x_2, \ldots, x_k$ represent the k possible values of attribute a. Then let $C(x_i)$ represent the number of chosen committee members with value x_i for attribute a. For categorical attributes, we compare the absolute difference $|C(x_i) - E(x_i)|$, where $E(x_i) = n/k$ for all x_i (i.e., each of the k attribute values in the committee of $n=10$ is equally likely). For each x_i, let $N(x_i)$ represent the absolute difference term, normalized between $[0,1]$. Then for categorical a, let balance $b_{diversity}$ be computed as follows.

$$b_{diversity} = \frac{1}{k}\sum_{i=1}^{k} N(x_i)$$

For numerical attributes (e.g., age and political experience), we define balance according to the number of even ranges of value covered by the committees. Let

x_{min} and x_{max} be the minimum and maximum values of attribute a, respectively. Let $d = (1/n)(x_{max} - x_{min})$, then define $n=10$ ranges as follows: $\{ [x_{min}, x_{min} + d), [x_{min} + d, x_{min} + 2d), \ldots, [x_{min} + 9d, x_{max}] \}$. Let C represent the count of each range for which a committee member has been chosen. Then, compute $b_{diversity}$ as follows.

$$b_{diversity} = \frac{n - C}{n}$$

Proportion: Proportion quantifies balance with respect to how much it proportionally represents the underlying distribution of the data. In spirit, it is similar to **AD**, but the computation differs slightly. It ranges from 0 to 1, where 0 represents a perfectly balanced committee, and 1 represents a highly imbalanced committee. According to the viewpoint that a balanced committee is one that is proportionally representative of the dataset, an ideal committee has people across possible values in proportion to their number in the underlying dataset. For example, a gender-balanced committee would have 1-2 women and 8-9 men (based on the underlying dataset's total of 13 / 100 women). An education-balanced committee would have 2 bachelor degrees, 2-3 master degrees, 3-4 law degrees, and 0-1 of each of the remaining degrees.

Balance from this perspective is computed the same a for $b_{diversity}$ with a few key differences. Categorical attributes for $b_{proportion}$ are computed in the same way, except that the expected values of each attribute $E(x_i)$ are based on the underlying distribution of the data, rather than equal numbers per value.

Likewise, balance for categorical attributes in $b_{proportion}$ is the same as for $b_{diversity}$ except that the ranges of numerical values are defined by equal deciles of the data. That is, assume all politicians P of the dataset are sorted by attribute a so that $p_1(a) \leq p_2(a) \leq \ldots \leq p_{100}(a)$. Then define $n=10$ ranges as follows: $\{ [p_1(a), p_{10}(a)], [p_{11}(a), p_{20}(a)], \ldots, [p_{91}(a), p_{100}(a)] \}$.

Ratio: Ratio is used to quantify balance for the “Gender” and “Political Party” attributes only, representing the percentage of “Male” or “Republican” politicians in the committee, respectively. While it ranges from 0 to 1 like the other balance computation approaches, **Ratio** does not imply good or bad based on the value itself. The interpretation is up to each individual.