1. Jaccard Similarity for Two Binary Vectors

1.1. Numerical Example

Say we are trying to compute the similarity of a store's customers. We could have a binary attribute that corresponds to an *item purchased* at the store, where 1 indicates that a specific item was purchased and 0 indicates that a product was not purchased.
Since there could be thousands of products in the store, the number of products **not** purchased by any customer is much higher than the number of products purchased. When computing the similarity between customers, we only want to factor in purchases of items. This means that the *item purchased* is an **asymmetric binary variable**, making a value of 1 more important than 0.
In the first step of a Jaccard Similarity measurement for two customers which consist of n binary attributes, the following four quantities (i.e., frequencies) are computed for the given binary data:

- a = the number of attributes that equal 1 for both objects i an j
- b = the number of attributes that equal 0 for object i but equal 1 for object j
- c = the number of attributes that equal 1 for object i but equal 0 for object j
- d = the number of attributes that equal 0 for both objects i and j.

Then, Jaccard Similarity for these attributes is calculated by the following equation:

$$J(i,j) = sim(i,j) = \frac{a}{a+b+c}$$

Notice the number of 0 matches is considered unimportant in this computation and is ignored because the items are asymmetric binary attributes.

Suppose that a customer transaction table contains 9 items and 3 customers. The similarity between objects is computed based only on the asymmetric attributes.

	item1	item2	item2	item4	item5	item6	item7	item8	item9
C1	0	1	0	0	0	1	0	0	1
C2	0	0	1	0	0	0	0	0	1
C3	1	1	0	0	0	1	0	0	0

The similarity between each pair of the three customers can be calculated by Jaccard Coefficient:

The similarity between each pair of the three customers can be calculated by Jaccard Coefficient:

$$J(C1, C2) = \frac{a}{a+b+c} = \frac{1}{1+1+2} = 0.25 J(C1, C3) = \frac{a}{a+b+c}$$

= 2 / 2+1+1 = 0.5