The Official Original Derivation of AQWV

This is the definitive official derivation of AQWV, which is really a modified version of ATWV.

Note that there are three main “input” weights: C, V, and an overall *a-priori* estimate for $P_{Relevant}$
The input weights do not directly include a cost for a miss.

Note well that there are two different kinds of $P_{Relevant}$ in this document.

- The first in an *a-priori* estimate for $P_{Relevant}$ **across** all the queries and all the datasets. This estimate enters into the factor called β (but **not** into the calculation of P_{miss} and P_{fa} for a specific system on a specific query on a specific dataset).
- The second is the actual $P_{Relevant}$ for a specific system for a specific query on a specific dataset. This actual statistic enters into the calculation of P_{miss} and P_{fa} (but not into the calculation of the factor called β).

This distinction will be clear by the end of the derivation of *Value*, below.

The input weights:

C is the cost of an incorrect or spurious detection (a false alarm)
(that is, returning a document that is not relevant)

V is the value of a correct detection
(that is, correctly returning a document that is actually relevant)

The value of $P_{Relevant}$ incorporated into β is an a-priori estimate of $P_{Relevant}$ across all the datasets, and is held constant acrosss the entire evaluations.

The variables in the formulas:

N_{fa} is the number of false alarms
(documents retrieved that are **not** relevant)
$N_{correct}$ is the number of correctly retrieved documents
(documents retrieved that **are** relevant)
N_{miss} is the number of relevant documents that were not retrieved

N_{total} is the total number of documents in the dataset
$N_{relevant}$ is the number of relevant documents in the dataset
$N_{nonRelevant}$ is the number of non-relevant documents in the dataset
Note that (Ntrue+NnonTarg) = Ntotal

$$P_{relevant} = (N_{relevant} \,/\, N_{total})$$

$$1/P_{relevant} = (N_{total} \,/\, N_{relevant})$$

We will use that formulation of $1/P_{relevant}$ in our derivation of *Value*, below.

$Value = [(V * N_{correct}) - (C * N_{fa})] / (V * N_{relevant})$ <== This is the original idea

$$= \frac{V * N_{correct}}{V * N_{relevant}} - \frac{(C/V) * N_{fa}}{V * N_{relevant}}$$

$$= \left(\frac{N_{correct}}{N_{relevant}}\right) - \frac{(C/V) * N_{fa}}{N_{relevant}}$$

$$= \left(\frac{N_{relevant} - N_{miss}}{N_{relevant}}\right) - \frac{(C/V) * N_{fa}}{N_{relevant}}$$

$$= \left(\frac{N_{relevant} - N_{miss}}{N_{relevant}}\right) - \left[\frac{(C/V) * N_{fa}}{1} * \frac{1}{N_{relevant}}\right]$$

$$= \left(\frac{N_{relevant} - N_{miss}}{N_{relevant}}\right) - \left[(C/V) * \frac{N_{fa}}{1} * \frac{1}{N_{relevant}}\right]$$

$$= \frac{N_{relevant} - N_{miss}}{N_{relevant}} - \left[(C/V) * \left(\frac{N_{fa}}{N_{total} - N_{relevant}}\right) * \left(\frac{N_{total} - N_{relevant}}{N_{relevant}}\right)\right]$$

$$= P_{correct} - \left[(C/V) * P_{fa} * ((N_{total} - N_{relevant}) / N_{relevant})\right]$$

$$= 1 - P_{miss} - \left[(C/V) * P_{fa} * (N_{nonRelevant} / N_{relevant})\right]$$

$$= 1 - P_{miss} - \left[(C/V) * (N_{nonRelevant} / N_{relevant}) * P_{fa} \right]$$

$$= 1 - P_{miss} - \left[(C/V) * \left(\frac{N_{nonRelevant}}{N_{relevant}} + \frac{N_{relevant}}{N_{relevant}} - \frac{N_{relevant}}{N_{relevant}}\right) * P_{fa}\right]$$

$$= 1 - P_{miss} - \left[(C/V) * \left(\frac{N_{nonRelevant} + N_{relevant}}{N_{relevant}} - \frac{N_{relevant}}{N_{relevant}}\right) * P_{fa}\right]$$

$$= 1 - P_{miss} - \left[(C/V) * \left(\frac{N_{total}}{N_{relevant}} - \frac{N_{relevant}}{N_{relevant}}\right) * P_{fa}\right]$$

$$= 1 - P_{miss} - \left[(C/V) * (1/P_{relevant} - 1) * P_{fa}\right]$$

If we let β denote $(C/V) * (1/P_{relevant} - 1)$, then we can rewrite the above as
$Value = 1 - P_{miss} - \beta * P_{fa}$ or as $Value = 1 - (P_{miss} + \beta * P_{fa})$

However, **note well**, the value of $P_{relevant}$ that we actually incorporate into β is not the value computed above. For β, we instead substitute our best *a-priori* estimate *across* our datasets, queries, and languages: we hold that estimate (and β) constant across entire evaluations so that we can make comparisons between systems and track progress from year to year. This allows apples-to-apples comparisons.

Value is for a particular query. The mean of *Value* over all the queries is QWV (Query Weighted Value). $AQWV$ (Actual Query Weighted Value) is QWV when the system is run at its actual decision threshold (we call that threshold θ).