

Transmail Relevance Formulae

There are three lists of attributes which could feasibly influence the Transmail's relevance to a receiver user. The three lists are groupings according to mail actions relevance, collaborative relevance and defined importance factors, which I shall label A , B and C respectively. The most obvious starting point is to work with indicator functions, $J(\cdot)$, which are simply "on" or "off", depending on whether the attribute is favourable or not. This is fine for attributes such as:

"If a user receives a TransMail and replies to it, then the content of the original TransMail is likely to be of significant relevance",

but not necessarily for attributes such as:

"If a user has read a given TransMail many times, then it is likely to be of significant relevance to them.",

where a score function may be more appropriate, with a higher score dependent on the number of times opened or read. However this causes problems of relative weighting amongst attributes, and in any case we can create an indicator function by defining a threshold value of number of times opened or read. The third list is different as it deals with importances (see later), and thus must have a continuous scale of values. These importance functions are called $I(\cdot)$, [and are the relative importances between users](#). A working formula for the relevance, $R(T_M)$, of a given transmail, T_M is then:

$$R(T_M) = \frac{1}{3} \left(\frac{1}{W_A} \sum_{j=1}^{n_A} w_j^A J(A_j) + \frac{1}{W_B} \sum_{j=1}^{n_B} w_j^B J(B_j) + \frac{1}{W_C} \sum_{j=1}^{n_C} w_j^C I(C_j) \right)$$

where n_A , n_B and n_C are the number of attributes in each list, w^A , w^B and w^C are weightings associated with each attribute to allow for the possibility that some attributes are deemed of greater intrinsic relevance than others, and W^A , W^B and W^C are the sum of the weightings in each of the three lists. Of course if the weightings are all unity in a list then $W = n$ for that list. This formula assumes that the relative maximum weighting of each list is equal to 1, a sensible option; it may also be sensible to weight each list in the ratio $W_A : W_B : W_C$, in which case the formula would be:

$$R(T_M) = \frac{1}{W} \left(\sum_{j=1}^{n_A} w_j^A J(A_j) + \sum_{j=1}^{n_B} w_j^B J(B_j) + \sum_{j=1}^{n_C} w_j^C I(C_j) \right)$$

where:

$$W = \sum_{j=1}^{n_A} w_j^A + \sum_{j=1}^{n_B} w_j^B + \sum_{j=1}^{n_C} w_j^C ,$$

the sum of the weight within and across the lists.

Importance

The third list is special in that it introduces the interactive aspect of relevance between users through entity importance, whereas the first two are purely independent in their individual factor contributions to relevance. This is because the intrinsic nature of these factors for mail actions by the user or by other users other than the user in question will affect the relevance of the Transmail, but the relevance of the Transmail will not feedback into the quantification of these factors. They either happen or they don't – the fact that the relevance may dictate whether these actions are engaged is irrelevant here, as we are merely revising relevance in the light of **past** actions. However, the importance factors are different. Here the values of these factors are affected by the email traffic that has occurred. For example the importance of a user is affected by the receipt of Transmails from other users – his/her importance perturbs or is shared by the importance of the other users. This interactive process must be taken into account before the importance can be fed into the equations (1) or (2) in a search. Similar arguments will apply for other importance factors, so we will assume that we can focus on one such importance factor, say the user importance factor.

To estimate the importance of a user the following model can be considered.

- 1) There is, for user i , an intrinsic immutable importance, I_i^0 , which is set globally and presumably takes into account levels of responsibility within the organisation. This intrinsic value should ensure that the mechanism doesn't result in total uniformity of user importance.
- 2) If a Transmail is sent from user i to user j , the importance of user j is amended slightly such that if user i is of higher importance than j then the importance of j is raised; however if the importance of i is lower than j then the importance of j is lowered. There is in essence a sharing of importance from the sender to the receiver. The importance of the sender is unaltered by the Transmail, but of course his/her importance **relative** to the receiver has been altered, such that the difference between them has been reduced.
- 3) The “sharing parameter”, α , dictates the degree of this sharing, with $0 < \alpha < 1$. In practice α will probably be high, say 0.9 or higher. There is also a parameter β to allow for the restraining to the preset level, I_j^0 .

With this, an equation to update user j 's importance at the k th update from Transmail's received since the last update, $k-1$ is:

$$I_j^{(k)} = \beta I_j^0 + (1 - \alpha - \beta) I_j^{(k-1)} + \frac{\alpha}{m} \sum_{i=1}^m I_i^{(k-1)} / C_i$$

$$I_j^{(k)} = \beta I_j^0 + (1 - \beta) \{ (1 - \alpha) I_j^{(k-1)} + \alpha \sum_{i=1}^m I_i^{(k)} / C_i \}$$

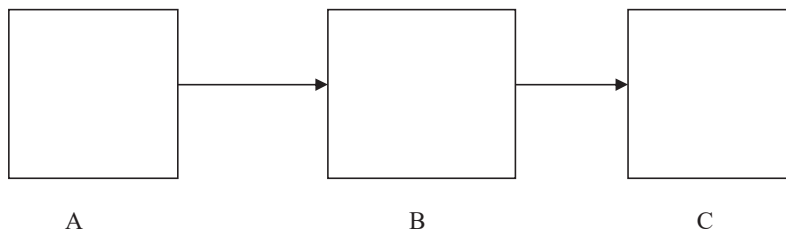
where C_i is the number of recipients of the i th email. Note in the summation the same user may appear more than once. This equation reflects two effects. Firstly we have a recourse back to a preset importance, I_j^0 , set by, eg, personnel criteria. This is governed by the mix, involving β , with the value of importance calculated from the second effect (in curly brackets). This second effect (in curly brackets) is the mix which accounts for adjustment due to received emails. This equation will be modified when a received email is a forwarded email (ie contains a history submessage), such that the importance of the forwarder i , is adjusted by that of the originator of the submessage, I :

$$I_i^{(k)} \rightarrow I_i^{(k)} + \alpha I_l^{(k)}$$

where C_i is the number of recipients of the i th email. Note in the summation the same user may appear more than once. Note unless we divide the third term in the formula above by m , we will always increase the importance of a user by receipt of Transmails. This should not, of course, be the case. Furthermore if no emails are received the importance is slightly reduced; this is feasible as it indicates a nonparticipatory aspect of the user. Note if no emails are sent there is a likely reduction in the importance using these equations. This may or may not be desirable, and can be removed by removing the $(1 - \alpha)$ factor if no emails are received.

The following examples were performed using simple coding of the equations (using Splus). In all 3 examples the preset values (I^0) were regarded as equal to the last update values, making the value of β irrelevant.

Example



User A has sent a Transmail to B who has sent a Transmail to C. Here:

$$I_A^k = \beta I_j^0 + (1 - \beta)(1 - \alpha) I_j^{k-1}$$

$$I_B^k = \beta I_B^0 + (1 - \beta) \{ (1 - \alpha) I_B^{k-1} + \alpha I_A^k \}$$

$$I_C^k = \beta I_C^0 + (1 - \beta) \{ (1 - \alpha) I_C^{k-1} + \alpha I_B^k \}$$

This system of equations ~~was will be~~ solved iteratively in what ~~was is~~ essentially a Gauss-Seidel algorithm procedure until convergence to a sensible precision. This procedure will then be repeated every regular time period, say daily. In this example, the more important A is, the more important B becomes, which then increases C's importance more than if B had not received the email from A.

For initial importances of A,B and C (ie I^0) of 1 and last update values of 0.7, alpha and beta both equal to 0.1, we reach convergence after 4 iterations to get importances for A, B and C of 0.667, 0.727 and 0.732 respectively. Note that A has reduced, whilst B and C have increased, but C has increased slightly more due to B's improved importance (as C recieved a Transmail from B).

Example2

As for example 1 except that the Transmail from B to C is a forwarded message of that from A to B. Here the three equations are:

$$I_A^k = \beta I_A^0 + (1 - \beta) (1 - \alpha) I_A^{k-1}$$

$$I_B^k = \beta I_B^0 + (1 - \beta) \{ (1 - \alpha) I_B^{k-1} + \alpha I_A^k \}$$

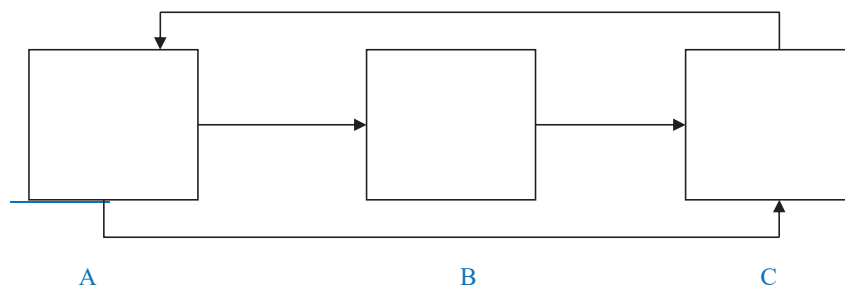
$$I_C^k = \beta I_C^0 + (1 - \beta) \{ (1 - \alpha) I_C^{k-1} + \alpha [I_B^k + \alpha I_A^k] \}$$

The only equation that is different is that for the imprtance in C which now reflects the forwarding of a message from A. Using the same input values, we now obtain 0.667, 0.727 and 0.738 for the updated importances for A, B and C respectively. The only alteration is that for C from the values in Example 1.

Formatted

Example 3

As for example 2 except that C now responds back to A, who then replies back to C:



Again using the same input values we arrive at updated importances of 0.740, 0.734 and 0.806 for A, B and C respectively.

The importance model above allows for relative importance for B with respect to A, as the adjustment resulting from a Transmail between A and B depends on the relative values for I_A and I_B . Ultimately it is these relative importances (the differences in importance between users) that are important in assessing relevance as used in the main relevance equation. These could be easily rescaled or adjusted to remove negative values in an appropriate way if so desired.

Importances calculated in this way will thus yield a square matrix of relative importance values of dimension equal to the number of users. A further issue here is that two senior people may have an equal high relative importance to a junior, yet one of these is in practice of little importance to the junior because he works in a different group. But by that consideration we are introducing the group entity, and thus we should now be considering the sum of the relative importances for the two entities (group and user).

Time-dependent Relevance

Returning to the main framework equation for relevance (for example equation (1)), we would like a received Transmail's relevance to generally a time-dependency. We should therefore premultiply the formula in equation (1) by a sensible function of time, $G(t)$:

$$R(T_M) = \frac{G(t)}{3} \left(\frac{1}{W_A} \sum_{j=1}^{n_A} w_j^A I(A_j) + \frac{1}{W_B} \sum_{j=1}^{n_B} w_j^B I(B_j) + \frac{1}{W_C} \sum_{j=1}^{n_C} w_j^C I(C_j) \right)$$

For example $G(t)$ might reflect relevance of a project up to a deadline time, t_0 , beyond which the relevance begins to decay:

$$G(t) = \begin{cases} 1 & t \leq t_0 \\ e^{-\lambda(t-t_0)} & t > t_0 \end{cases}$$

where λ is a decay constant (high for rapid decay, low for slow decay).

Notes

1) These formulae are additive in their nature between the three lists, but for importance is interactive within. A further degree of sophistication might be to include degrees of effect, say on user action which is dependent on the importance of the user or other entities. These are then interactions between the actions and the importances.

2) There is no effect on the sender's importance of the Transmail in this model. We could model alternatively with the adjustment to importance being made to the sender rather than the receiver, or both.

3) The model for importance is reminiscent of that used for the PageRank system by Google, but is not exactly the same. The same iterative procedure, which will employ the basic Gauss-Seidel algorithm is employed however.

4) One sophistication not catered for here is that discussed in section 5.4 in that (relative) importance could be different depending on the direction of the Transmail. In some ways this is a function thereof of the action rather than the user, and may be catered for in an interactive model with interactions between importances and user actions.

EXHIBIT 2047

metadata slip sheet

BegDoc	REV0000099
BegAttach	REV0000099
File Name	TransmailRelevanceFormula1.doc
LastCreateOn	8/10/2004 10:08 AM
LastModifiedDate	9/9/2004 10:33 AM
Custodian	Marston, Justin
Author	Morton-Jones
SortDate	9/9/2004 10:33 AM
Office-Revision Number	38
Last Saved By	Morton-Jones
Exhibit No	