9.1 The Development of the Mathematical Algorithms

9.1.1 Introduction

The delivery of the TransMail system of allocating relevance and importance will require the continuous development of a highly complex mathematically based system. There are some similarities in this to the Google development of page-rank technology.[1] Algorithms will be required that continually update a set of inter-related values, based on user action, so that everyone's actions affect everyone else. There are, however, also significant differences from the Google model, in particular:

- Each user/post will have a different individual set of importance and relevance criteria
- TransMail also requires the input of user-set values for the relevance and importance criteria, which both correct the values for the user concerned and also offer a check on the effectiveness of the particular algorithms being used.

This present section sets out the basic mathematical framework within which the detailed mathematical model will be developed.

9.1.2 Aim of Relevance Formula

The key aim of the relevance system is to produce two values:

(1) The relevance ranking of an incoming TransMail before it is opened. $R^I(T_m)$, *its incoming relevance.*

(2) The relevance ranking of a TransMail after it is opened and read. $R^P(T_m)$ *its post-reading relevance.*

This is not exactly the same as incoming and stored relevance, because some TransMails may be stored without being opened. In fact there could be identified three different levels: (1) Relevance on arrival before being seen (2) Relevance after being opened and read (3) Relevance after been seen but dealt with unopened. This, however, will be ignored for this initial model.

The ultimate aim of the system can be expressed as follows:

> "If the recipient were to open a set of incoming mails and assign relevance scores and rankings, these would correlate highly with both $R^I(T_m)$ and $R^P(T_m)$."

9.1.3 Factors in Relevance Formula

There are three lists of attributes that influence the Transmail's relevance to a receiver user.

The first are indicator functions, D_j^m, which are dichotomous, ie are simply "on" or "off", depending on whether the attribute is "favourable" or not. This includes attributes such as:

General Factors (available for both $R^I(T_m)$ and $R^P(T_m)$:

[1] See http://www.webworkshop.net/pagerank.html and also http://ausweb.scu.edu.au/aw2k/papers/lowe/paper.html

"Flags for follow up – the sender indicates the TransMail requires action by the recipient".

Factors available only for $R^P(T_m)$

"The user replies to the TransMail" *[If so then D=100 if not then D=0]*

A second set of variables C_k^m are values on an ordinal or even a cardinal scale, eg:

General Factors (available for both $R^I(T_m)$ and $R^P(T_m)$:

"the ratio of others who received it to the number who opened it"

"the priority = the urgency tag given it"

Factors available only for $R^P(T_m)$

"how many times the user reads the TransMail"

"to how many people the user forwards the TransMail"

These are placed into suitable functions $f(C_k^m)$ which are between 0 and 100.

Within the general factors referred to above, in an instance in which this is the first TransMail received from this source by a particular user, the concept of Mail Actions Collaborative Relevance is particularly useful. This concerns any actions done on this TransMail by others whom the recipient considers to have high local importance (see also below),

Some of these $f(C_k^m)$ may also "decay" over time. It may be difficult to determine this without some indication from the user, but obviously there are some TransMails connected with projects that, once completed, leave the user unlikely to wish to reconsult the TransMails. Some feedback may be useful connected with a particular job or project code.

There is also a third factor in relevance, one that deals with the importance of the sender or forwarder. This is a complex concept in that there is both an absolute global importance but also an importance to the particular recipient. A recipient's line-manager may have low importance in the great scheme of things, but high importance to the recipient him/herself. This means that there are two factors here:

I_g^m is the global importance of the sender

I_l^m is the "local" importance of the sender (ie importance to that individual recipient)

N^m is the number of recipients of the TransMail

Importance is again on a scale of numbers, and is put into a function eg $f(I_l^m)$ which is between 0 and 100.

There are also importance factors involved in cases where the immediate sender is forwarding mail from a previous sender, who also has his/her own importance values. We can indicate this by I_l^{om} to indicate any earlier forwarder(s) of the TransMail.

A summary of the working formula for the relevance, $R(T_M)$, of a given transmail, T_M is then:

$$R(T_m) = \frac{\omega}{W_D}\sum_{i=1}^{n_I} w_i D_i^m + \frac{\upsilon}{W_C}\sum_{j=1}^{n_J} w_j f(C_k^m) + \frac{\psi}{W_I}\left((w_1 f(I_g^m) + w_2 f(I_l^m) + w_3 \sum_{k=1}^{n_K}\left[w_1 f(I_g^m / N^m) + w_2 f(I_{lk}^{o,m} / N^m) \right] \right)$$

where

n_i n_j and n_k are the number of attributes in each list

w_i and w_j and w_1 and w_2 are weightings associated with each attribute

W_D, W_C etc are the sum of the weightings in each of the three lists

$\omega + \upsilon + \psi = 1$

As it stands the formula allows the separation out of the three types of input. If preferred, then the weights can be suitably readjusted and the formula can be rewritten in a simpler form.

With appropriate values for the weights, then, the relevance of piece of mail m is a number between 0 and 100.

Local importance is generally likely to be more crucial to relevance of incoming TransMail than global importance. Even if a TransMail is received from someone very high up in importance (eg a Company Director) it is likely to have been sent to many recipients. The relevance formula is therefore a function of the importance divided by number of recipients as indicated

There may be occasions when the relevance of incoming TransMail is highly affected by the life cycle of a project. This would mean that the relevance could be multiplied by a time-related factor *G(t)*. For example $G(t)$ might reflect relevance of a project up to a deadline time, t_0, beyond which the relevance begins to decay:

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

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

Most stored TransMail will have a time-decay-based element of relevance.

9.1.4 Relevance Feedback loop

As noted, a sample is collected by the system of the actual relevance scores that a user would have assigned, having looked at the TransMail, on a sliding scale which effectively works as a percentage. A “good” system would produce *a priori* relevance weightings that turned out to be near to the ones assigned post-opening by the recipient.

The formula used above is actually a linear combination of the various given values of the D, C and I ‘**s.** It is therefore possible, given enough feedback data, to use linear regression to derive the weights which would minimise the error-squared terms in predicting the scores for relevance as given by the user. The data base used is essentially a very large matrix, with a large number of zeros and is often likely to be ill-conditioned. Specialised mathematics is needed to deal with such situations, and internal tests for absurdity (which can be thrown up when linear regression matrices are ill-conditioned) inbuilt.

This analysis enables two things. First, different recipients who feedback could have their own individual weighting estimated. Secondly, if patterns of weightings emerge for different types or levels or importances of user, then global best estimates of weightings on a user category basis can be refined

over time and these will be useful particularly in assigning initial weighting to a new user.

9.1.5 Importance

The factors labelled D_j^m are things that either happen or don't happen eg "does the recipient reply?" and C_k^m known measurements eg "to how many users does the recipient forward it?" Their values are not going to be readjusted in the light of the success or otherwise of the prediction of relevance values for incoming TransMails – only the *weightings* given to them will be adjusted using the linear regression model as above.

The importance factors I_g^m and I_l^m are different, because they are not clearly ascertainable variables but are (estimated) parameters constructed within the model itself. Here the values of these factors are affected by the TransMail 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.

There is, for user i, some true global value I_{gi}^{Ω} to which ideally the system should converge. In practice, of course this will never be static (importances are constantly changing), but it is a useful idealisation. Initially a starting value of I_{gi}^{0} will be set globally, based on judgements of levels of responsibility within the organisation. The ideal is that starting from I_{gi}^{0} the system will converge over time to a value near to I_{gi}^{Ω}.

It is also necessary to add in a factor that enables importance to be carried over from posts. If user-A moved from post-S to take over post-T, then user-A's global importance will be some function of their previous importance whilst in post-S, and the importance the previous holder of post-T had. Likewise local importance will also be a function of post as well as individual. This has been omitted from the mathematical description below for simplicity at this stage.

Importance attaches to the sender, but also to the job code and context, which are also (passive) entities within the system. The job-code type carries an importance value alongside the sender. This added complexity means that actually the relevance formula has to contain weightings of both types of importance – sender and job code type and perhaps some other factors.

9.1.6 Factors affecting global importances over time

(1) **<u>TransMails received:</u>** Receipt of an TransMail from a more important sender increases the importance of recipient – dependent on how many others receive the same TransMail. "Shared importance" therefore depends on ${}^{r}I_i/N_i$ where N_i is the number of recipients of the i th TransMail. The effects depend both on the number of TransMails and the average of the shared importances.

The factor affecting the global importance at time k of a user is given by:

$$f\left(\sum_{i=1}^{m} {}^{r}I_i^{k-1} / N_i \,;\, \frac{1}{m}\sum_{i=1}^{m} {}^{r}I_i^{k-1} / N_i\right)$$

(summed across the m senders in time k-1 to k)

(2) **Results of TransMails sent**: A user whose TransMails are never opened or forwarded will decrease in importance, whereas one whose TransMails are usually read becomes more important [this may also stop people sending unnecessary TransMails!]. If those of greater importance read and deal with TransMails this can also increase the user's importance – unless it is always the same recipient (as would be the case of a PA TransMailing his/her boss regularly). Safeguards concerning the latter may be reviewed, though in some instances a PA acting for a more important person may need a higher importance rating if they regularly act on that person's behalf.

Again the effect of this factor is a function of both the number and average importance of the recipients and the results. Importance sharing of those who opened and read the TransMail is a function also of the time spent with it open, relative to the length of the TransMail itself:

Those who opened it: $$f\left(\sum_{i=1}^{m} {}^{r}I_i^{k-1}(t_i / L_i); \frac{1}{m}\sum_{i=1}^{m} {}^{r}I_i^{k-1}(t_i / L_i)\right)$$

where t_i is time spent and L_i a function of its length (obviously an important one-line TransMail can be read in twenty seconds) giving an "attention ratio" of t_i/L_i

For those who read it, there is also an effect depending on whether they:

(i) replied to it

(ii) deleted it without reply (this could have a negative effect)

(iii) stored it

(iv) forwarded it.

This can be indicated by an A representing action (and various other actions are included as described in section 5.3.1-2 above):

$$f\left(\sum_{i=1}^{m} {}^{r}I_i^{k-1}(A); \frac{1}{m}\sum_{i=1}^{m} {}^{r}I_i^{k-1}(A)\right)$$

This can be countered by a negative effect on importance of recipients who either deleted (del) it or stored it without reading it (no).

$$-f\left(\sum_{i=1}^{m} {}^{no}I_i^{k-1}; \frac{1}{m}\sum_{i=1}^{m} {}^{no}I_i^{k-1}; \sum_{i=1}^{m} {}^{del}I_i^{k-1}; \frac{1}{m}\sum_{i=1}^{m} {}^{del}I_i^{k-1}\right)$$

(3) **Relevance Rankings:** If recipients give received TransMails a high relevance score (when asked on the sample basis), then this will increase the importance of the sender.

If relevance reported by recipient is $R^r(T_i)$ for a transmail from the user, then the importance is affected by a function of the average of the n such reports made during the time period k-1 to k, ie: :

$$f\left(\frac{1}{n}\sum_{i=1}^{n} f'({}^{r}I_i^{k-1}; R^r(T_i))\right)$$

This also allows feedback to the system to ensure that it converges to some kind of "realistic" estimates. Direct feedback on "importance" may be very difficult to build in because most people will be likely to have a better idea of importance *to them* than global importance. Some kind of periodic review eg from HR (at least dealing with in-house relative importances) could be built in however as the system develops.

A global importance formula will readjust the global importance in time k from that of k-1 using these various functions:. The value α indicates how fast the global importance should change over time.

$$
\begin{aligned}
I_{gi}^{k} = \alpha I_{gi}^{k-1} + (1-\alpha)\Big[& \mathrm{w}_1 f\Big(\sum_{i=1}^{m} {}^{r}I_i^{k-1} / N_i \,;\, \frac{1}{m}\sum_{i=1}^{m} {}^{r}I_i^{k-1} / N_i\Big) \\
& + \mathrm{w}_2 f\Big(\sum_{i=1}^{m} {}^{r}I_i^{k-1}(t_i / L_i);\, \frac{1}{m}\sum_{i=1}^{m} {}^{r}I_i^{k-1}(t_i / L_i)\Big) \\
& + \mathrm{w}_3 f\Big(\sum_{i=1}^{m} {}^{r}I_i^{k-1}(A) \,;\, \frac{1}{m}\sum_{i=1}^{m} {}^{r}I_i^{k-1}(A)\Big) \\
& - \mathrm{w}_4 f\Big(\sum_{i=1}^{m} {}^{no}I_i^{k-1} \,;\, \frac{1}{m}\sum_{i=1}^{m} {}^{no}I_i^{k-1} \,;\, \sum_{i=1}^{m} {}^{del}I_i^{k-1} \,;\, \frac{1}{m}\sum_{i=1}^{m} {}^{del}I_i^{k-1}\Big) \\
& + \mathrm{w}_5 f\Big(\frac{1}{n}\sum_{i=1}^{n} f'({}^{r}I_i^{k-1}; R^{r}(T_i))\Big)\Big]
\end{aligned}
$$

These global importances will be recomputed by the system every period at intervals as determined by the operators ie by bluespace. They are, of course, heavily inter-related and so this system of equations is solved iteratively in what is essentially a Gauss-Seidel algorithm procedure until convergence to a sensible precision. As the system is developed there are more powerful iterative systems that can be applied.

9.1.7 Factors affecting local Importance over time

This is a comparatively more straightforward concept, at least for TransMails received from previous contacts. It may be preferable simply to recompute this each time (or used an updated stored recomputation) rather than work by adjusting a starting value. This will be based on a time-dependent function similar to that used in exponential smoothing, so that more recent actions have greater effect.

Relative importance will be affected, of course, by global importance I_{gi}^{k-1}, but also by:

(1) <u>Results of Past Received TransMails</u>:

This is basically actions based on past receipts. These should be weighted, based on a time dependent function as mentioned, so that more recent actions have greater weighting.

i. The number/percentage of times previous TransMails have been deleted or stored without opening.

$$\sum_{q=1}^{n} f_q^{t}(del; no)$$

here f^t is a time-based function for TransMail q, which is a function of whether the TransMail is deleted or stored unopened.

ii. The 'attention ratio' of t_i/L_i for TransMails opened, and number of times.

$$\sum_{q=1}^{n} f_q^t(t_{i1}/L_{i1};\, t_{i2}/L_{i2\cdots})$$

[Other actions as listed in section 5.3.1-2 are also included].

(2) <u>TransMails sent:</u>

This is a function, time related, of how many times the recipient has previously sent "unsolicited" TransMails to this sender"

$$\sum_{p=1}^{n} f_p^t$$

(3) <u>Relevance Rankings</u>:

Any past relevance scores given by the recipient to TransMails from this sender." (a weighted average if more than one has been graded).

$$\sum_{r=1}^{n} f_r^t(R(T_r))$$

(4) <u>Mail Actions Collaborative Relevance/Importance</u>

When a new contact communicates with a user for the first time, the system can still make a 'best guess' about that new contact's importance, and a key element of this is collaborative importance. Suppose that user-A receives a TransMail from user-B. This is the first such TransMail, but user-A regards user-C and user-D as having high local importance, and both of these have now ranked user-B as locally important. This will increase the initial "best guess" of local importance the system ascribes to user-A for user-B. This implies a formula as follows:

$$\sum_{c=1}^{n} f(I_l^{c\ for\ sender};\, I_l^{resipient\ for\ c})$$

(5) <u>User Importance Determination</u>

Within the window when a TransMail is opened will be a slider showing the local importance of the sender as calculated by the system. The recipient will be able to adjust this if (s)he feels that it is inappropriately low or high. If this is the first time a TransMail has been received from this sender, then the recipient can opt for a setting which will flash to remind the him or her that (s)he has not previously received any TransMail from this source, and that readjusment of the calculated importance of the sender might therefore be particularly useful.

The last set value from the recipient would be $I^{r\,t}_l$

This means that the overall figure for local importance is determined thus:

$$I_l^m = \alpha_t I^{r\,t}_l + (1-\alpha_t)\Big[w_1 I_{gi}^{k-1} + w_2 \sum_{c=1}^{n} f(I_l^{c\ for\ sender};\, I_l^{resipient\ for\ c}) + w_3 \sum_{q=1}^{n} f_q^t(del;no)$$

$$+ w_4 \sum_{q=1}^{n} f_q^t(t_{i1}/L_{i1}; t_{i2}/L_{i2} \ldots)$$

$$+ w_5 \sum_{p=1}^{n} f_p^t$$

$$+ w_6 \sum_{r=1}^{n} f_r^t(R(T_r)) \quad]$$

If there are no values for $I^{r\,t}_{l}$ then the α_t value applies is zero.

If no previous TransMails have been received from or sent to this sender, then revised weightings will apply to the first two elements. If the sender is unknown to everyone in the "collaborative group" then a "first guess" can only be made with reference to the global importance.

In general there is less interactiveness involved in this local importance formula, and only the second term above becomes fluid. This means that, if another member in the "collaborative" and mutually locally important group acts in such a way as to increase the importance of a relatively unknown sender, then this could affect the priority given a TransMail in the in-box of a recipient. In simple terms, if user-A has a TransMail from user-B waiting unread, but that user-A's collaborators act so that the local importance of B is increased (eg by actions done on user-B's transmails to them) the this could increase the assigned local importance of user-B to user-A, and so increase the relevance of the waiting TransMail from user-B in user-A's inbox.

9.1.8 Content Analysis

As this system develops and computer sophistication increases, the intention is to add some elements of content analysis. This will involve the system itself using spiders to identify terms within the content of any TransMail which statistically increase the likelihood that it is relevant to the recipient.

9.1.9 Policing the System

As this system develops, users may seek to find ways to increase their own global importance. Thus eg two friends on similar grades could arrange to send each other large numbers of TransMails which they would leave open for a while – each therefore increasing the other's importance. This is not a "prisoner's dilemma" game-theory situation because each can monitor the other's actions over time, and large numbers of read TransMails could increase both their importances. The advantage of this to them is that both would then get their TransMails read faster when sending to others. In another scenario, when the content analysis is advanced, senders could deliberately introduce phrases and words to get their TransMails read (although this might annoy recipients and so have a negative effect on their importance ranking in the long run).

All this is analogous to those who work out (in some cases employing whole teams of analysts) how to move their company or group up the Google ratings. The bluespace software engineers will be working on this to incorporate detection devices identifying this, and to penalise it where and as appropriate.

9.1.10 The Possibility of Serious Error

It must be remembered that this is fundamentally an aid, and is never going to be a perfect model. Someone coming from outside the TransMail system, into a new post, who has never before TransMailed anyone or worked on any jobs with them, can have an "importance" guestimated only perhaps on a basis of

some input from a HR department. It is likely that such a person may in practice need to communicate in other ways to ensure that their TransMails are not simply ignored. This, of course, presently happens anyway with emails – busy people tend to simply delete or ignore emails from unknown and unrecognisable senders because most of them are probably spam anyway. If an email filtering system is used to see only emails from known senders then this will have an even more obvious effect. The system is a valuable aid, but not an infallible one.

EXHIBIT 2005

metadata slip sheet

BegDoc	REV0000093
BegAttach	REV0000093
File Name	Pauls Formula for Justins Business.doc
LastCreateOn	9/20/2004 4:01 PM
LastModifiedDate	5/12/2005 2:12 AM
Custodian	Marston, Justin
Author	.
SortDate	5/12/2005 2:12 AM