

時間的表現に基づく優先度判定モデル

5Q-7

加来田 裕和 角 隆一 山本隆広 玉井詩子
NTT ソフトウェア研究所†

kakuda@slab.ntt.co.jp

1 はじめに

インターネット上には WWW や NEWS, メーリングリストなど様々な情報メディアが発達してきている。また、これら情報メディアを利用したオンライン新聞などの商用サービスが増えつつあり、欲しい情報を容易に手に入れられるようになってきた。これら情報メディアから、必要な情報を効率よく取得するためには、情報の優先度を判断しフィルタリングを行う必要がある。

従来、個別メディアに対して、利用者自ら嗜好に合せて情報をフィルタリングしたり、利用者が年齢、性別やキーワードを登録することによってフィルタリングしたりする方法があった。より精度の高いフィルタリングを行う場合は、情報の内容に踏み込んだ理解をすることが必要である。しかし、自動で内容を完全に理解し判断することは困難である。

2 情報の優先順位

筆者は、実際の電子メールより優先順位を判断する言葉を抽出した。これら言葉を、以下のように分類した。ここでいう優先順位とは情報を処理する順番のことを表している。

●期日

「○月○日」や「×曜日」、「本日」などの表現

●緊迫した状態を表す表現

「急いで」や「早急に」、「回答願います」などの表現

●情報発信者

人名やメールアドレスなどの表現

●情報の内容を表すキーワード

「釣り情報」や「コンサート情報」などの表現

●情報到着日

情報が到着した日付

ここで、「情報の発信者」「情報の内容を表すキーワード」は、興味に関する項目と考える。つまり、興味のある情報を選択しようとしたとき、内容による判断だけでなく情報の発信者が誰であるかも特定していると考え、この2つをあわせて処理順位を算出し

たものを興味度とする。また、「期日」、「緊迫した状態を表す表現」、「情報到着日」を時間に関係する項目であると考え、これらは、急いで処理しなければならない情報を選択しようとした場合、判断のためよく用いる。この3つの項目をあわせて処理順序を算出したものを緊迫度とする。

3 優先度判定モデル

精度よくフィルタリングを行うには、以上の項目から情報の優先順位を判断しなければならない。そこで、情報の優先度を判断するモデルとして図1のモデルを提案する。

これは、優先度を興味度と緊迫度から、下記のように計算することにより、情報へ優先順位付けする。

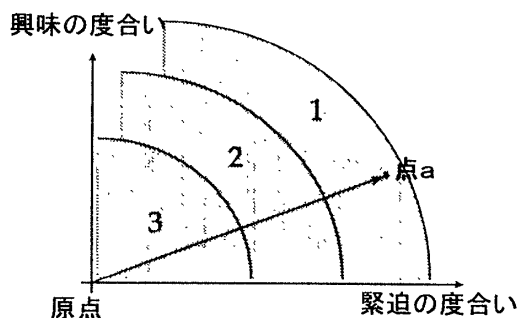


図1 優先度判定モデル

3.1 興味度モデル

$$(\text{優先度})^2 = (\text{興味度})^2 + (\text{緊迫度合})^2$$

興味度モデルは、情報から情報発信者や内容を表すキーワードを抽出して、この情報に対する興味度を示すモデルである。これを、以下の式で表す。

情報の発信者は、受信者との組織上の関係など人間関係によって重み付けする。また、キーワード

$$(\text{興味度}) = a(\text{発信者}) + b(\text{キーワード})$$

は、業務上の関係や趣味など関係の深い事柄によって重み付けする。

3.2 緊迫度モデル

緊迫度モデルは、情報から日時、緊迫の表現な

†Timely Sorting Model for Information Filtering
Hirokazu KAKUDA Ryuichi SUMI Takahiro YAMAMOTO
Utako TAMAI
NTT Software Laboratories

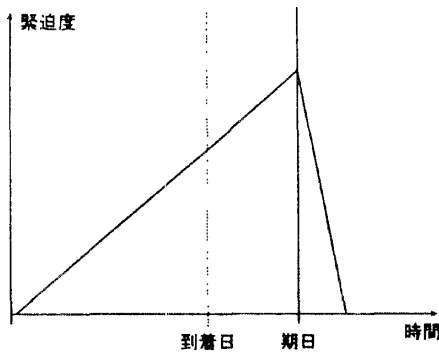


図 2 緊迫度推移の例

などを抽出することによって、その情報に対する期日前後の緊迫した状態を示すモデルである(図 2)。これを以下の式で表す。

記述である期日や緊迫した状態の表現を抽出することによって、情報の緊迫度が計算でき、情報への優先順位付けが可能となる。

$$(緊迫度) = c(期日と到着日) + d(緊迫表現)$$

たとえば、「その日時まで」という期日付きの情報の場合には、その日時に近づくほど情報の重要さが増す。その日時を超えると重要さが減少する。また、「その日時から」という情報も、同様である。

4 評価

優先度判定モデルの評価のため、実際のユーザサポートメールへ優先順位付けを行った。

優先順位付けは、以下の3種類を行った。

1. 興味度のみによる順位付け
2. 緊迫度のみによる順位付け
3. 優先度(興味度と緊迫度)による順位付け

これらと、筆者自身が主観的にメールを順位付けしたものとを比較した結果、表 1 が得られた。

数値は、全メールのうちその順序が付けられたメール数である。優先順位は、「もっとも高い」「高い」「普通」「低い」「もっとも低い」の 5 段階である。

優先順位	緊迫度	興味度	優先度	筆者の主観
もっとも高い	3	4	3	4
高い	12	22	24	20
普通	19	8	10	13
低い	5	16	13	15
もっとも低い	21	10	10	8

※全メール数:60 通

表 1

このうち、主観によって順位付けしたメールが、それぞれの度合いによって順位付けしたメールのなか何割合含まれているか示したものが表 2 である。

優先順位	緊迫度	興味度	優先度
もっとも高い	75%	75%	75%
高い	55%	75%	82%
普通	77%	38%	60%
低い	13%	73%	87%
もっとも低い	100%	100%	100%

表 2

5 考察

表 2 より、主観によって順位付けしたものと優先度で順位付けしたものを比較すると、この優先度は精度よく順位を判断できていると考える。また、緊迫度と興味度をそれぞれ適用した場合と、優先度を適用した場合を比較すると、特に「高い」と「低い」とした順位に着目すれば、優先度を適用した場合の方が、精度よく優先順位を判断できている。

期日や緊迫している表現から緊迫度を算出しただけ、また興味や人名から興味度を算出しただけより、優先度としてすべての項目を判断し順序付けすることが必要であることが、表 2 で示されたと考える。

6 まとめ

優先度を判定する項目により順位を付加する場合、その付加する値をいかに決めるかが重要になる。

本評価では、興味度、緊迫度の式中 a,b,c,d というパラメータは 1 としていた。このパラメータの値も重要である。今後は、様々な種類の情報に対して、期日、人名、組織、情報から緊迫していることを表す言葉を抽出して優先順位付けを行い、言葉に対する値付けの方法やパラメータの値を決定する。さらに優先度判定モデルを評価していく。

参考文献

- [1] 森田:情報フィルタリングシステム, 情報処理, 37(8), pp.751-758(Aug. 1996)
- [2] Paul Resnick: "Group Lens: An Open Architecture for Collaborative Filtering of NetNews", In Proceedings of CSCW, pp. 175-186 (1994)