情報距離管理方式

2F−7

佐藤　基
NTT通信網総合研究所

1.はじめに

近年、オフィス内文書の電子化、高速ネットワークによる情報共有が急速に進められている。ワープロ、電子ファイリング、電子メール、電子掲示板等、個々に管理されていた情報がネットワークを介して共有された将来、固定的に分類困難な大量の情報が発生する。これらの情報を扱うための方法として、ファジィシソーラスによる動的リンク[1]や、動的／非定型／断片的な情報を管理するノウハウ蓄積システム[2]などがあるが、いずれもキーワード等の情報に対する定義が強要される。

情報のあいまいな管理／検索を実現するため、ユーザの興味に基づく情報間の関連性を情報間の距離として記述することにより、情報のあいまいな分類／検索を可能とする適応情報ネットワークシステム（ANT:Adaptive information NeTwork system）の研究を行なっている。本稿では、情報距離管理、特に情報間の距離を検索条件として利用する場合に問題となる情報間の最短距離問題について報告する。

2.適応情報ネットワークシステム

ANTは、複数の端末(100~1000台)がネットワークを介してファイル共有可能な環境において、膨大な情報のなかから希望する情報を入手するため”興味ある情報を近づけ、興味のない情報を遠ざける”機構を提供する。具体的には、情報間の関係の度合いを情報距離という数値で表現し、情報の書き込みや検索等の情報操作により抽出された情報への興味をもとに情報距離を更新し、検索条件として距離を指定することによりデータ検索やブラウジング範囲の制限を行なう。これにより、自分が興味を持った情報を指定し、近傍に存在する情報を入手可能とする新たな情報探索手段が提供される。以降、ユーザの興味という抽象概念が、情報間の距離に正しくマッピングされたと仮定して議論を進める。

情報距離を検索条件とした場合の情報探索では、基点となる情報に直接関連づけられている情報が近くにあるとは限らない点に特徴がある。図1では、情報Aから直接関係づけられた情報Dよりも、3つの情報をへて間接的に関係づけられた情報Gの方が近いと判断される。情報間の距離を検索条件とする際に、最も近いルートのみをリストする機構が必要となる。例えば図1で情報Aから距離3以内を指定した場合に、情報Cが2度リストされることを避けなければならず、最短距離のみが有効となる。この問題を情報間の最短距離問題と呼ぶこととする。

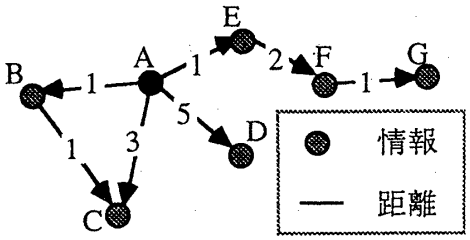

図1　情報距離のイメージ

3.情報間の最短距離問題

情報間の最短距離問題は、情報をノード、距離をエッジと置き換えることにより”重み付き有向グラフ”と考えることができる。情報間の最短距離問題の特徴を以下に示す。

＜情報間の最短距離問題の特徴＞

（1）ノード数が多い

（2）エッジ数が少ない疎なグラフである

（3）エッジの重みは、ネットワーク上での物理的な距離とまったく無関係である。

（4）情報検索時に正確な最短距離を必要とするため、近似を利用できない。

（5）ノードおよびエッジの追加／削除頻度が高い

（6）エッジの重みが小さい（近い距離）場合の利用がほとんどである

情報の記憶方法としては、（1）および（2）よりマトリクスによる記憶ではなく、自情報を起点とする距離を情報ごとに記憶する隣接リスト管理を行なう[3]。また、（3）および（4）から、ネットワークにおけるルーティング問題のような階層管理による近似を利用できず、最短距離を正確に計算することが前提となる。

A study of information distance control
Motoi Sato
NTT Telecommunication Network Laboratories

4.最短距離の計算方式

情報の書き込み時など距離が更新されるごとに最短距離を計算する最短距離記憶方式と、検索時に計算する逐次計算方式について比較する。

4.1 最短距離記憶方式

最短距離記憶方式は、情報の書き込み時などの情報間の距離が更新されるごとに最短距離を計算し、各情報ごとに他情報への最短距離を記憶しておく方式である。すべてのノードに対する最短経路を求めるアルゴリズムとして、DijkstraとFloydのアルゴリズムが利用される。n個の情報とe本の距離をもつ場合、Floydのアルゴリズムの計算時間のオーダが$O(n^3)$であるのに対し、隣接リストを利用したDijkstraのアルゴリズムではO(en log n) となり、3項(2)よりDijkstra のアルゴリズムを採用する。

4.2 逐次計算方式

逐次計算方式は、検索が実行された時点で、基点となる情報から指定された距離以内で到達可能な情報への最短距離を計算する。1つの情報に関する最短距離を求める方法としてDijkstraのアルゴリズムが適用でき、n個の情報とe本の距離をもつ場合、計算時間のオーダはO(e log n)となる。Dijkstraのアルゴリズムでは、距離が近い情報から確定してゆくことから、検索時に指定された距離およびリストされる情報数により制限される。リストする情報数をaに制限すると、計算時間のオーダはO(a log n)となる。

4.3 方式比較

表1に最短距離記憶方式と逐次計算方式を比較する。一般に、情報更新に比べて情報検索を行なう数が多いことから、システムトータルの計算時間は最短距離記憶方式の方が少ない。しかし、2項(5)より距離更新がネットワーク上に分散された複数の端末より同時に発生し易く競合管理が困難なこと、ユーザが情報の更新時に待たされることに慣れていないことを考慮すると、検索時に計算し1件あたりO(a log n)という比較的少ない時間で計算可能な逐次計算方式がマンマシン上有利となる。

5.まとめ

ユーザの情報への興味を尺度とする情報間の関係を、情報間の距離として表現する適応情報ネットワークシステムにおける情報距離管理方式について示した。情報距離を検索条件として指定する際に情報間の最短距離問題が発生し、これを計算するための方式として最短距離記憶方式と逐次計算方式を提案し、比較を行なった。両方式のいずれがマンマシン上有利であるかという結論は、プロトタイプの利用実験により検証する。

[参考文献]
[1] 小川, 森田, 金矢, "ハイパーテキストのためのファジィ動的リンク機能", 人工知能学会研究会資, HICG-9101-2, pp.9-18 (1991)
[2] 関, "ノウハウ蓄積支援システムの構築", 信学技報, OS-91-4 (1991)
[3] 石畑, "アルゴリズムとデータ構造", 岩波書店, pp.223-276 (1989)

表1.最短距離計算方式の比較

n：情報数, e： 距離, a：検索リスト数

項目＼方式	最短距離記憶方式	逐次検索方式
情報書込時間	書込時間 + O(en log n)	書込時間
情報検索時間	検索時間	検索時間 + O(a log n)
総計算時間	距離更新回数 x O(en log n)	検索回数 x O(a log n)
記憶容量	無限大を除くすべての最短距離を情報ごとに記憶する	定常的に記憶するものはない
その他	複数の距離更新が発生した際の競合制御問題がある	距離更新がない場合でも、常に最短距離を計算しなければならない