

大量なグラフデータを対象とする可視化ツール

渡邊都明† 塚本享治†

東京工科大学メディア学部メディア学科

1 はじめに

関係をもつデータ群を視覚化することで、データを観察・分析・捜査・監視することが可能になる。しかし一般的なシステムは、それぞれのデータことに対応させた可視化システムになっていることが多い。汎用的な可視化システムを開発中であるので、報告する。

2 データの可視化

2.1 データの構造

データフォーマットには様々な形式があり、類似的なシステムの多くはシステム毎に独自のフォーマットを用いていることが多い。[1][2]しかし個別のデータ構造では汎用的に扱うことができないので、共通化するためのデータ構造を定義することにした。二つ組のデータノードと、ノード間を結ぶ関係であるエッジによる、三つ組のデータを基本のデータ構図に採用することで、データとデータの関係を示すことができ、ネットワーク構造を表現することが可能であると考えたからである。(図1)

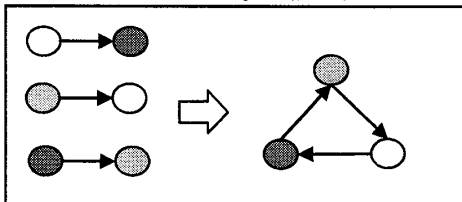


図1 データ構造

2.2 大量のデータを扱う可視化システム

2.2.1 リソースの問題

データをネットワーク越しに分散している場合、もしくは巨大なデータを扱う場合、一度に全てを取得するのはメモリやネットワークに対して、不必要な負担をもたらす。この問題を考慮して、データを取得する処理を分散させる必要がある。

2.2.2 視覚的な問題

画面に大量なデータをそのまま映し出すことが、意味を持つとは限らないし、余分なデータを表示してしまうことは、時に害悪である。また単語単位のような検索による絞り込みでは、大量のデータに対して優れたレスポンスを返す

ことができなかつたり、関係性を示す必要がある。そこでデータが構造上、隣接しているデータとだけ関係性を持っていることを考慮して、焦点となるノードからの距離による絞り込みを行うことにした。(図2)

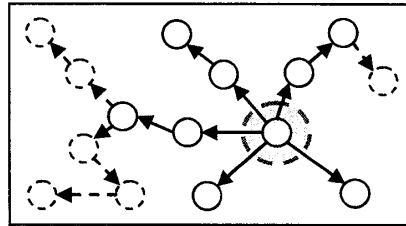


図2 データの絞り込み

2.2.3 解決としてのウォークスルー

各問題点は、データの構造を、道を歩き進んで行く人物の視点で表現することで解決を試みた。ノードを選択すると、選択したノードに中心が移り変わり、新たな中心のノードから隣接伝いの距離に応じて、画面に映し出されるノードが導き出される、ウォークスルー形式である。

3 システムの実現

3.1 システムの全体像

システムは、三つ組のデータを二つの球と直線によって表現する。ノードやエッジを表現するのに際し、拡大縮小や回転、中心から幾つの距離まで表示する等といった調整をするための操作は、タブやスライダーといった一般的な GUI によって行う。ノードの位置を手動で仕分けする場合や、中心のノードを選択する場合に関しては、マウスによる操作がインターフェースとして適切であると判断した。また隣接するノード数が多い場合に、特定の関係をピックアップすることができるようにした。(図3)

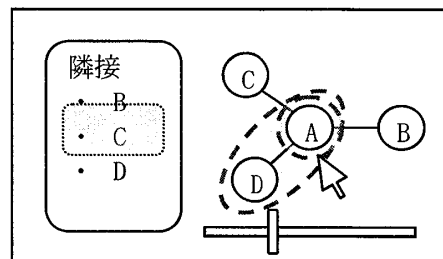


図3 インターフェースの設計

Visualization for a quantity of graph data

†Kuniaki Watanabe, Michiharu Tsukamoto

†Tokyo University of Technology School of Media Science

3.2 データの設計とトラバースの設計

データベースには、CSV や XML など様々なデータ形式があるが、大量のデータに対する検索の高速性を重要視した結果、RDB の形式に限定することにした。O/R マッピングを用いて変換することで、本システムとバインドする。

オブジェクトへ変換されて、システム内に表示されたデータのノードを選択すると、中心が移り変わる。中心が移り変わったことに対するイベント駆動に対応して、中心から設定された距離までの周囲のノードを検出することで、データのトラバースを実現した。

3.3 画面の構成

画面には、以下の四つの役割を持たせた。

- (1) データを可視化して描くキャンバス
- (2) 表示形式を制御するコントローラ
- (3) 隣接するノード一覧を表示するパネル
- (4) データファイルを指定するコントローラ

それぞれの役割によってウィンドウを一定の割合で分割した領域として振り分ける。(4)に関しては、メニューバーとしてファイル選択ダイアログボックスを出力する仕組みとなっている。まとめると、以下の図のようになる。(図 4)

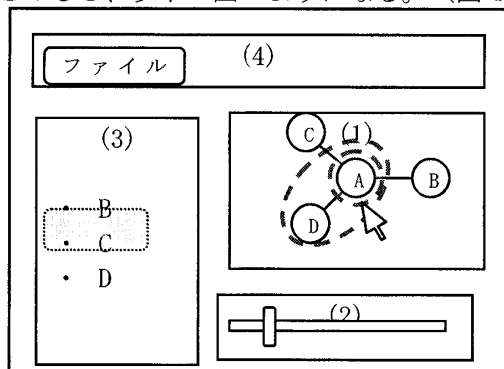


図 4 画面の設計

3.4 表示方法の操作

表示方法は、以下のように編集できる。

- (1) 画面の拡大と縮小、回転
- (2) 中心ノードを画面の中心に固定する
- (3) ノードを動き回らせる
- (4) 表示するノードの中心からの距離
- (5) ラベルを表示する

中心ノードを画面の中心に固定すると、隣接ノードを伝って移動する際に見やすい。

表示されるノードの数が膨大であると、画面内でノードが重なって描画されてしまうことがあるので、一定の範囲で動き回らせている。

しかしこれらの機能の中には、大量に表示されたノードを観察するとき、またマウスで操

作するとき邪魔になることもあるので、停止させられるようにしてある。

4 評価と課題

4.1 評価

類似したツールの多くは、ノード間の関係を単純な線や矢印によってのみ表すことができるが、本ツールでは、ER 図やクラス図で扱われているような、様々な関係線を示すことができる点で優れている。

4.2 課題

(1) システムの高速化

単体のデータベーステーブルに、全てのデータを単純に入れているので、データの検査効率が悪い。またグラフを画面に描画する際において、グラフィックの高速化が必要である。

(2) 外部ソフトウェアとの連携

本稿のツールでは、関係を示すことができるが、各ノードが何を示すのかについて表現することができない。しかし本システムに詳細な処理を直接に組み込むと、システムの拡張性を限定してしまう。そこで本システムにサーバーとしての役割をもたせて、外部ソフトウェアと連携することで、用途に幅を持たせたい。

(3) 多種多様なデータへの適用

本稿のツールでは、扱うデータを予め高速に扱える形式で、従って RDB で用意しておく必要があった。しかし多種多様なデータを、予め手動で RDB に変換するのは困難である。標準のテキストフォーマットである XML 形式のデータから、システムが高速に扱うためのデータ形式へと変換する補助的なシステムが必要である。

5 おわりに

本稿では、インターフェースに関する基礎として、グラフを生成するツールを開発した。本システムは、様々なシステムを開発するための基礎となり、そのための応用性を有している。今後はシステムの拡張を通して、視覚的なインターフェースから、ツールとしてのインターフェースへと発展させたい。

6 参考文献

[1] 菱田 哲史 「動的ネットワークの変化過程の可視化手法」筑波大学卒業論文(2008)

[2] 美馬 秀樹 「MIMA Search」

<http://ciee.t.utokyo.ac.jp/MimaSearch/index.html> (2010/1/1)