

第 XIX 部

ネットワーク統計情報の可視化技術

第19部 ネットワーク統計情報の可視化技術

第1章 netviz wg について

netviz ワーキンググループの目的は、ネットワークの視覚化およびその他の表現、そのためのツールやノウハウに関する議論と情報の共有を行なう事である。

現在、WIDE には、ビジュアライゼーション自体を研究テーマにしている人はいないが、目に見えないインターネット技術を説明するため、また、説得力のあるプレゼンテーションのために、道具としてビジュアライゼーションを使いこなす事はすべての研究者に必要なになっている。従来は、mawi ワーキンググループなどで計測データの視覚化等の議論をしてきたが、研究目的を持ったワーキンググループの中では、どうしても研究内容に話が行くため、参加者の間口が狭くなる。そこで、独立したワーキンググループとして、データ表現にフォーカスした netviz で活動を行っている。

第2章 Twitter サービスの可視化

2.1 Twitter サービスの可視化

今年は身近な題材の可視化を行い、可視化の過程やツールなどについての共有を行なった。

近年、140文字の「つぶやき」を投稿し、共有する Twitter[176] という Web サービスが注目を集めている。WIDE Project のメンバーの中にも実際にサービスに登録し、「つぶやき」を投稿しているメンバーもいる。

また、近年の Web サービスでは API を利用して簡単にデータにアクセスする手段を用意している。Twitter もこの例に漏れず、API を公開し、http/https 経由でデータへのアクセスが可能となっている。

そこで、本年は身近な題材として Twitter をとり

あげ、可視化のテーマとした。

本稿では既存の Twitter の可視化サービスについて調査をおこない、実際に Twitter の可視化を行なった経緯を報告する。

2.1.1 2010年5月 WIDE 研究会

2010年5月28日から29日に名古屋大学で開催された WIDE 研究会において、twitter のデータを可視化した既存プロジェクトの調査結果について報告した。

ここで報告したサービスは「つぶやき」とユーザを関連づけて可視化するようなサービスが主流であることがわかった。また、Twitter では様々な種類の「つぶやき」が投稿されるという性質を反映してか、Twitter 上での流行を把握するためのサービスも目立った。

2.1.2 2010年9月 WIDE 合宿

5月の WIDE 研究会での発表を受けて、実際に Twitter のデータを使用して可視化を試みた。

2010年9月8日から11日に信州松代ロイヤルホテルで開催された WIDE 合宿において、試行錯誤の過程を含め、Twitter の可視化した内容について報告をおこなった。

2.1.2.1 可視化の目的

Twitter を利用している WIDE Project のメンバーが投稿した「つぶやき」から似た傾向をもつユーザの可視化ができないか試行錯誤をおこなった。

2.1.2.2 取得したデータ

Twitter を利用している WIDE Project のメンバーを特定するにあたり、WIDE_Project (https://twitter.com/#!/WIDE_Project) というアカウントに注目した。このアカウントは非公式ではあるが、WIDE Project のメンバーが取得し、WIDE メンバーが「つぶやき」を投稿できるようにしている。このアカウントがフォローしているアカウントが WIDE Project のメンバーであろうと予測し、公開されている「つぶやき」のみを取得した。

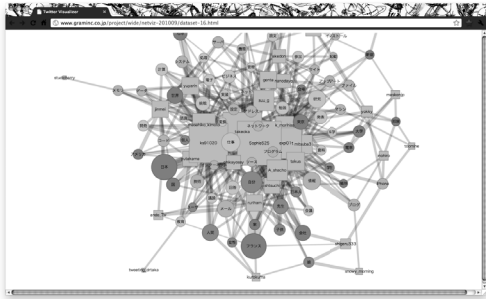


図 2.1. Force-Directed アルゴリズムを用いた可視化の 1 つ

また、aonoe が作成した wide というリスト (<https://twitter.com/#!/aonoe/wide>) でフォローされたアカウントも WIDE Project のメンバーであろうと予測し、公開されている「つぶやき」のみを取得した。

2.1.2.3 可視化の内容

取得した「つぶやき」を日本語形態素解析機の mecab[126] を利用して、単語のみを抽出した。抽出した単語とユーザの関連性を可視化をおこなった。可視化を行なうにあたり、3つの手法を試みた。

また、aonoe の wide リストでフォローされていたアカウントは数が多かったため、可視化用データを生成する時間が足りず、WIDE_Project がフォローしていたアカウントのデータのみを使用した。

まず、protovis という javascript で使用できる可視化ライブラリで利用できる Force-Directed アルゴリズムを使用した可視化を行なった (図 2.1)。

この可視化ではアカウントを四角のノードで、「つぶやき」に含まれる単語を円のノードで表現している。また、各ノードの大きさは可視化で示されている単語をよく投稿しているほど大きくなり、よくつぶやかれた単語ほど大きく表示している。

また、単語は適当にグループ分けをおこない、似た内容の単語を同じ色で表示するようにしている。

単語を通して似た傾向の「つぶやき」を投稿したアカウントが分かりやすいようにすることが目的であったが、各ノードの数が多すぎてノード間のつながりが分かりにくくなってしまった。各ノードをマウスでクリックしたら関連するノードの色が変わるなどの仕組みが必要である。

また、雑然として描画されている印象が強いのでアカウントのノードを周囲に円形で配置し、円の中

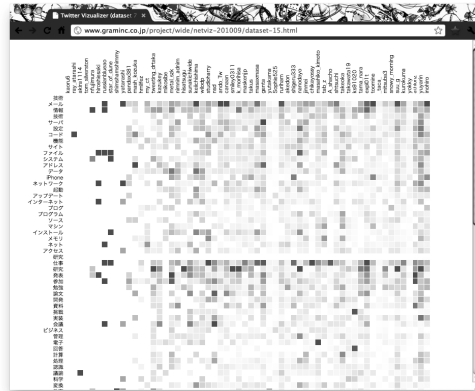


図 2.2. マトリックス状に配置した可視化の 1 つ



図 2.3. タグクラウド風の可視化

に単語のノードを表示するなどの工夫が必要である。

次につぶやかれた単語とアカウントをマトリックス状に配置し、よくつぶやかれた組み合わせを目立つように色分けをおこなった (図 2.2)。

この可視化では各要素が多すぎるため、一覧性にかけてという欠点があった。また、つぶやきの総数が少ないアカウントや単語は必然的に薄い色になってしまうため、データの正規化などが必要である。

また、マトリックスのある要素をクリックするなどの動作で要素の情報を表示できるような仕組みが必要である。

最後に各単語を月ごとに集計し、上位 20 位の単語についてタグクラウド状に表示した (図 2.3)。

各単語にマウスを当てると単語をつぶやいたアカウントの上位 5 位までをホバー表示するようにした。これまでの可視化と異なり時間軸の概念を導入したため、月ごとによくつぶやかれる単語が変化し、世間の話題とリンクした可視化の可能性を示せた。(例として Apple 社の新型 iPhone の発売の前後に合わ

せて「iPhone」という単語が上位に表示されるようになった。)

ただし、単語単体では意味不明になってしまうケースもあり、単語の選出方法などに課題が残った。

2.1.3 まとめ

本稿では Twitter が公開しているデータの可視化の内容について報告を行なった。今日では Twitter の他にも Web を通じて利用するサービスでは API 経由でのデータの取得方法を提供しているサービスは多い。このため、本報告の内容は Twitter に限らず他の Web サービスの可視化への応用も可能であると考えられる。

ただし、実際に作成した可視化の結果は必ずしも最初の目的を満たしたとは言いがたく、可視化の手法、とくにデータの集計方法に課題が残った。また、可視化のツールとして使用したライブラリの理解も十分とは言えず、結果的に伝えたいことが伝わるような可視化ができなかった。

可視化の目的に応じた収集したデータの集計方法などについて引き続き調査したい。

第3章 まとめ

インターネット研究において、計測データをいかに表現するかは重要なテーマである。しかしながら、WIDE にはデータ表現を主テーマとするような研究者がいないこともあり、なかなか WG が活性化しない状況が続いている。いっぽうで、興味を持つ研究者は多いので、活動を活発化する工夫をして行くつもりである。