

6L-04 学校間交流支援のための発信 Web コンテンツの分析

鈴木 雅実†

村松 茂樹†

松本 一則†

橋本 和夫†

†通信・放送機構

†KDD 研究所

1 学校間交流の支援

最近、インターネットの教育利用のための環境整備が進みつつあり、種々の高速な教育用のネットワークを介した情報交流が期待されている。ただ、現状では、先進的な少数の学校を除いて、Web ページの立ち上げや教育への活用方法の検討が端緒についた段階と言える。

文部省・郵政両省の主導による、「学校における複合アクセス網活用型インターネットに関する研究開発」(学校インターネット)においても、各導入地域の研究課題の主要項目として、学校(地域)間の交流が挙げられている。学校間交流に期待される要素としては、次のような項目が指摘できる。

- 遠隔地域の学校との間でコミュニケーションを体験することにより、生徒に生き生きとした学習素材を提供することができる。
- 各生徒が生活している地域の特色について他地域と比較することにより、価値観の多様性等を学ぶことができる。

一方、実際にこのような交流学習を実行する上での問題点として指摘されているような点は、表 1 のような事項である(参考文献 [1] による)。

表 1: 交流・共同の問題点
(2 項目選択, 計 200%で表示)

指摘のあった問題点	割合
1 メールアカウントが不足している	46%
2 国内交流・共同学習の相手が見つからない	26%
3 児童・生徒のプライバシーが保てるか不安	23%
4 教育効果がうまく評価できない	21%
(以下省略)	

効果的な交流方法の開拓も重要ではあるが、その以前に、表 1 の 2 項に対応する、交流相手を選択する際に役立つ情報の提供が求められている。そこで、その一手段として、各学校から発信される Web コンテン

ツをドキュメントの集合とみなし、コンテンツ内の語の出現頻度等の分析に基づいて学校毎の特徴付けを行ない、視覚的なサイトマップに投影する方法を検討した。すなわち、各学校の特徴が Web ページに反映されているものと考え、その特徴を可能な限り明確な切り口で捉えようとするものである。

2 Web コンテンツ分析とサイトの特徴表現

学校間交流の相手を探す際には、地域内/地域間での各学校の特色を多角的に把握することが望ましい。そこで、自校を含む地域の学校群の情報発信の特徴や、地域間での相違の把握を助ける情報の見取図としてサイトマップを構築する。ここで言うサイトマップとは、学校からの情報発信の傾向を示す特徴ベクトルにより各学校の特徴付けを行ない、それらを視覚化したものである。以下に、最初の試行として行なったサイトマップ作成手順、および実データのサンプルから見た現状の問題点を記す。

2.1 サイトマップ作成手順

1) 学校からの発信 Web コンテンツの収集

分析対象とする各学校毎の Web コンテンツ集合(サイト・ドキュメント)をネットワーク上から収集する。

2) ドキュメント間の類似度計算

サイト・ドキュメント内の各ドキュメント内の自立語の出現頻度分布に基づく、確率型モデルの文書間類似度計算を行なう [2]。

3) 全ドキュメント集合のセグメントへの分割

前記の類似度計算により、個々のドキュメントをノードする(ルートを持たない)木構造を最小スパン木として構成した後、全体を幾つかのセグメント(部分木)に分割する。その各セグメントを 1 つの次元と考え、個々のドキュメントは自身が属するサイト内外のドキュメントとの類似度に応じて、各セグメントに分散した形で配置される。その分布状況が、サイト全体の特徴を表わすことになる(セグメントへの分割基準については、文献 [3] を参照)。

4) 特徴ベクトルによるサイトの特徴表現

分割されたセグメント(次元)によって表現される空間上で、サイト別の各セグメントに属するドキュメントの頻度分布をベクトル形式で表現したものをサイト特徴量と呼ぶ。

5) 視覚的なサイトマップの構成

前述のサイト特徴量に基づいて、主成分分析を用いた2次元の視覚化を行ない、サイトマップとして表現する。図1にその例を示す。この例では、ある地域に属する33校の小学校からの発信コンテンツ（サイト・ドキュメント）を分析した結果を、2次元のサイトマップとして表現されている。なお、各学校サイト（S1～S33）の平均ドキュメント数は約28であった。

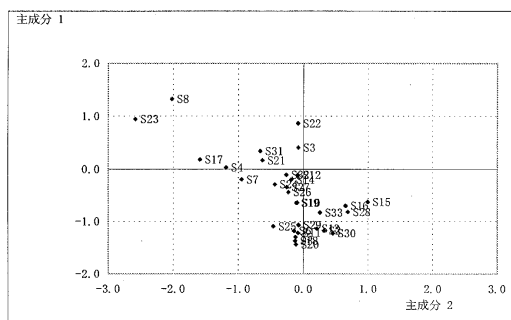


図1: 学校サイトマップの例 (同一地域内の小学校)

2.2 現状の問題点

図1では、実際には、主成分1および2の軸のプラス方向に、大きく離れた位置に、学校S9および学校S1がそれぞれプロットされている（図は両校を除いた中心部分を拡大したものに相当）。この2校についてはドキュメント数が多く、しかもある特定の話題領域に関するものが比較的集中していることが観察される。これが、2校の他の学校からの相対的な距離を遠くしていると言えるが、ドキュメント数の分布の偏りがある程度勘案しつつ、次節に述べるように、サイトの特徴を表現するための前提条件や人間の直観とも適合するような比較手法・尺度を再考する必要がある。

3 今後の課題

3.1 サイトマップ表現方法の改良

今回述べた手法では、サイト・ドキュメント内のドキュメント集合をすべて同列に比較している点で、やや問題がある。分析対象とするドキュメントを何らかの基準に基づいて事前に絞り込んでおくことや、サイトを特徴付ける要素としてドキュメント内の語の分布に注目する現手法以外の着眼点の検討も必要である。

また、現段階では、このサイトマップで示される各学校の特徴について、主成分の意味（ラベル）付け等を自動的に行なうまでには至っていないが、利用者への有効な情報提供手段として、検討を要する問題である。一方、観点を換えることにより、サイトマップ全体の

分布状況が変化するような多様なビューの提供も、視覚的な特徴表現の技法として有効と考えられる。

さらに、本手法で用いている類似度計算等の高速化や、サイトの特徴を表わすのに適したセグメント分割手法の検討が必要である [3]。

3.2 関連研究との比較

本研究では、各学校サイトの特徴を表わすサイトマップを、視覚的な情報発見等に役立てることをねらいとしているが、類似する研究事例として、動的に変化するサイト・ドキュメントの特徴を時系列的に表現したり、サイトを利用するユーザのプロファイルや要求に応じて、サイト内の情報を可視化する機能を持つサイト・アウトライニングがある [4]。

一方、個人レベルでの協調学習/作業を支援するため、個々の学習者が対話を通じて問題解決のためのコミュニケーションを行なう際の対話履歴から抽出したキーワード情報に基づいて、学習者の動的なグループ編成を可視化するシステムの提案もなされている [5]。

本研究では、サイト・ドキュメントの動的な変化の側面については現段階では重視しないが、視覚的な表現方法の効果等について、これらの先行事例を参考にしながら、目的に適した方式を探究することとしたい。

4 まとめと謝辞

本研究でねらいとしている点は、広い意味での特定の「話題」に関わる発信情報の背景にある、サイト（すなわち情報発信者/集団）の特徴そのものを、適当な切り口から可視化することにより、サイト同士の比較や交流相手としての選別を支援することにある。このため、Webコンテンツの分析および特徴の表現方法を、教育利用の視点から、より多くのデータを用いて改良・検証を進める予定である。

なお、本研究は、通信・放送機構（TAO）の直轄研究「学校における複合アクセス網活用型インターネットに関する研究開発」の一環として実施しているものである。関係各位の支援と助言に感謝いたします。

参考文献

- [1] 越桐 國雄：“インターネットの教育利用の現状 '98.1”，<http://www.osaka-kyoiku.ac.jp/educ/enq98/>, 1998.
- [2] Makoto IWAYAMA, Takenobu TOKUNAGA: “Hierarchical Bayesian Clustering for Automatic Text Classification”, Proceedings of IJCAI-95, pp.1322-1327, 1995.
- [3] 村松・鈴木・松本・橋本：“MDL基準を用いた文書集合の特徴化手法”，2000年電子情報通信学会総合大会，2000.
- [4] 武田・中村・浦本：“XMLがもたらす創造的ネットワーク—動的な情報源と分散エージェント—”，人工知能学会誌，Vol.14, No.6, 1999.
- [5] 齋藤・安達他：“協調学習における動的グループ編成の可視化”，電子情報通信学会1999年ソサイエティ大会，D-15-3, 1999.