

MANACO による地域観光情報観測手法の提案 Network Analysis for Regional Tourism with MANACO

堀井 洋十 林 正治† 堀井 美里‡ 吉田 武稔†

Hiroshi HORII Masaharu HAYASHI Misato HORII Taketoshi YOSHIDA

1. はじめに

昨今、地域の歴史・文化・伝統に注目した地域観光が脚光を浴びている。平成 18 年度版観光白書[1]によれば、平成 17 年度における国民 1 人当たりの国内宿泊観光旅行回数は 1.82 回と推計され、対前年度比で 6.4%増となっている。近年の地域観光では、インターネットの普及に伴い観光情報の発信および取得がネット上で行われることが一般的となっており、観光地（ホスト）側にとっては、観光客および観光予定客（ゲスト）が対象地域に関してどのような嗜好・興味を有しているのかを正確に把握することが非常に重要である。

一方、著者らは、メタデータ照合型ネットワーク解析システム MANACO(Metadata-based Analyzer for Network Appearances and Communications)の開発に取り組んでいる[2]。MANACO はネットワーク経路上において直接的に通信パケットを取得し、それを基にネットワークトラフィックが有する意味的特徴とその動向の解析を目的としたネットワーク解析システムである。通常のネットワークトラフィック観測では、中継通信機器からの提供情報や計算機上に蓄積された通信ログを情報源とするのに対して、本システムではネットワークケーブルから Layer1TAP 装置により電氣的に複製・分離した通信パケットを利用する。そして、取得した通信パケットを復元・整形した後、形態素解析処理を行うことでネットワークトラフィックに含まれる特定の語句を抽出・分類する。さらに MANACO では、パケット取得処理や取得情報の形態素解析処理などの主要処理機構において、オープンソースソフトウェアとして標準化されたソフトウェア技術を基盤として採用している。

本研究では、インターネット上において発信・取得されている地域観光情報の観測および解析手法として、MANACO を用いた地域観光情報観測手法を提案する。MANACO を利用することにより、観光において最も基本的かつ重要なホスト（観光客）・ゲスト（観光地）の概念を観光情報観測に反映させることが可能となり、歴史・文化などの意味的特徴の把握が実現される[3]。本論文では、MANACO を用いた観光情報観測の基礎的な仕組みと、地域観光における意味的特徴の把握例として、石川地域を対象にした観光情報の史学的特徴把握について論じる。本論文の構成は、以下の通りである。第 2 章において、観光情報の観測手法について記す。第 3 章では、著者らが提案する MANACO について、第 4 章では観光情報の観測と石川県歴史観光に対する応用例について記す。第 5 章はまとめである。

2. 従来研究

従来から、ゲスト側の興味を把握する観光情報観測手法として、検索エンジンの利用および Web サイトにおけるアクセスログの利用が行われてきた。Google[4]に代表される商用検索エンジンでは、検索ロボットと称される探索プログラムを使用して Web コンテンツ内容を収集（クロール）・蓄積する。Google で用いられている PageRank アルゴリズム[5]は Page らによって提案されたアルゴリズムであり、Web のリンク構造のみを用いて Web ページ群から客観的な人気度が高いと思われるページを機械的に算出することを目的としている。PageRank アルゴリズムでは、「大多数の良質なコンテンツを含む Web ページからリンクが張られている Web ページは、良質な Web ページと判断できる」という前提に基づいた再帰的な関係を利用して、関連する全ての Web ページの重要度を判定する。このような検索エンジンの特性に着目した観光情報解析に関する研究が行われている。中戸[6]らは、利用者の興味があるトピックを多く含む地域情報コンテンツを優先的に提供する地域情報検索システムを提案した。一方、アクセスログを利用した手法として、加地屋ら[7]は、北海道の道路情報総合案内サイトのアクセスログを利用した利用者ニーズ分析を行った。この研究では、ホスト側である Web サイトの 6 年間のアクセスログに基づいて利用者のニーズ分析や情報の重要度の推定を行い、その結果をコンテンツ内容に反映させる試みを行っている。

しかし、商用検索エンジンを利用した観光情報観測では、得られた観測結果に対して観光におけるゲスト・ホストの概念を適用することは困難である。これは検索ロボットによるクロール対象が動作時点で公開されている Web コンテンツに限定されており、特定地域を対象にした利用者の閲覧行動は観測対象に含まれていないことに起因する。また、Web サイトに蓄積されたアクセスログを利用する手法では、対象地域外からのアクセスを IP アドレス等から判別することにより、ゲスト・ホストの関係性を把握することは可能であるが、個々の Web サイト毎にアクセスログが蓄積されるため、対象地域（ホスト側）全体を対象にした観光情報観測は困難である。

本研究では、ホスト（観光地）およびゲスト（観光客）という観光において基礎的な概念・関係を反映した観光情報観測手法を提案する。対象地域を中心としたホスト・ゲストの関係性を明らかにするためには、観測対象地域および地点を明確にし、観測した結果から観光に関する意味的な情報を抽出することが不可欠である。これに対して MANACO は、ネットワーク物理層から直接的にネットワークトラフィックの観測を行い、通信が有する意味的な情報の把握が可能であることから、本研究では観光情報分野に対して MANACO の適用を試みる。

†北陸先端科学技術大学院大学 知識科学研究科

‡金沢大学大学院 社会環境科学研究科

	SNMP/NetFlowの利用	アクセスログの利用	Layer1TAP装置の利用
観測対象	通信機器・サーバー	サーバー	ネットワーク通信路
情報の種類・形態	主として時系列数値情報	文字情報	ネットワークバイト情報
観測・被観測媒体の関係	非同一	同一	非同一・不可視
観測データの再現性	高い	高い	低い(可読処理が必要)
観測データ形式の多様性	低い	高い	可読処理手法に依存
観測データの安全性	通信機器・プロトコルの安全性に依存	不正改竄の実例が多数報告	高い(被観測側からの接続は不可能)

表1 ネットワークトラフィック観測手法の比較

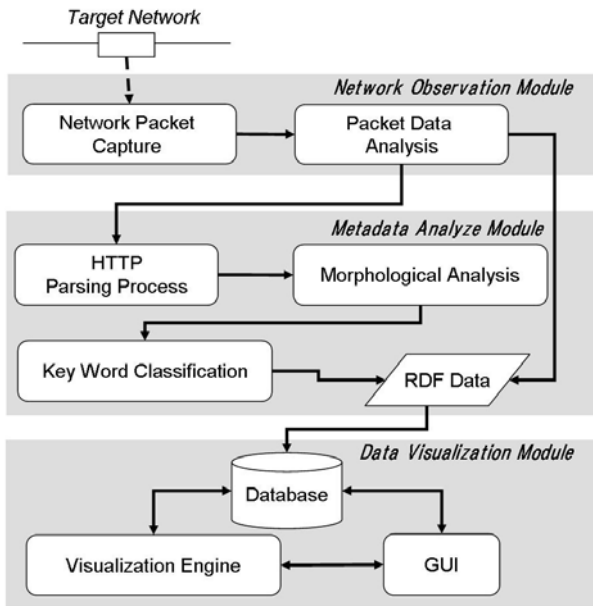


図1 MANACO システム処理構成

3. MANACO (Metadata-based Analyzer for Network Appearances and Communications)

本研究で提案する MANACO は、ネットワーク通信路から Layer1TAP 装置を介して直接的にネットワークパケットを取得し、ネットワークトラフィックに含まれる意味情報をリアルタイムに抽出・分析および可視化することを目的としたシステムである。提案システムの動作概要を図1に示す。現在、使用目的および用途に応じた様々な形態のネットワーク装置が提案・開発されている。しかしながら、その多くは OSI プロトコル層(Layer 7)までを対象としており、本研究の特徴である通信の目的および意味そのものに着目・特化したネットワーク装置および観測技術は著者が調査した限りにおいて提案・実用化されていない。本システムはネットワークパケット取得・意味的特徴解析・観測情報可視化の3基本モジュールにより構成する。図1に MANACO のシステム処理構成を示す。

一般的なネットワークトラフィックを対象にした観測では、通信機器から SNMP(Simple Network Management Protocol)[8]等を利用して通信機器の動作情報を取得する手法と、サーバーに蓄積されたアクセスログを解析する手法が広く一般的に用いられてきた。これに対して MANACO では、Layer1TAP 装置を利用してネットワーク通信路から直接的にネットワークパケットを取得し、その解析結果から意味情報の抽出を行う。Layer1TAP 装置はネットワーク

通信路から電氣的に信号を分離・複製するネットワーク機器であり、観測対象のネットワークトラフィックに対して速度低下や欠損などの影響を与えない、被観測ネットワーク側から Layer1TAP 装置の存在を特定・検出することは論理的に不可能であるなどのステルス性を有する。これら Layer1TAP 装置が有する特長を活かした提案手法と従来ネットワークトラフィック観測手法について、ネットワークトラフィック観測で重要な各項目の比較を表1に示す。

MANACO によるネットワークトラフィック観測の特長は、以下の2点である。

通信路からの直接的な通信パケット取得：現在、多くの情報ネットワークでは、VPN(Virtual Private Network)などの仮想化技術や動的構成技術が普及しており、ネットワーク構成・トポロジーについて論理・物理間の乖離が進んでいる。論理的ネットワーク構成に基づいたネットワークトラフィック観測を行う場合には、構成の仮想化・冗長化および実際の物理構成の差異を常に意識する必要があり、ネットワークトラフィック観測の煩雑さや信頼性に影響を与えていた。これに対して提案手法では、Layer1TAP 装置を用いたネットワークの物理層からの直接的なネットワークパケット取得および観測を行うため、複雑かつ動的に変化する論理的ネットワーク構成およびその変化に依存しない。

観測動作のステルス性：本手法が有するステルス性は、「観測活動が本来業務に対して影響を与えない」という観点からインターネットにおいては、非常に重要な要素の1つである。また、インターネット上の悪意を持った攻撃者から観測システムの存在そのものが不可視であることは、観測システム上に蓄積された観測情報に対して行われる不正改ざんやネットワーク攻撃の危険性が低減され、観光情報の分析において必須である信頼性・真正性の高いネットワークトラフィック観測が実現できる。

4. MANACO を用いた意味的特徴の把握

MANACO による観光情報ネットワークトラフィックからの意味的特徴把握の適用例として、本研究では歴史観光資産が数多く現存し、全国的にも知名度が高い石川地域を対象に史学的意味の抽出を試みる。図2に示す観測概要を示す。本提案では、首都圏などの石川地域外に居住し石川地域を将来の観光先としている入域予定客をゲストとし、石川地域内から観光情報コンテンツを制作・発信し観光客誘致を目指す観光企業・自治体をホストとする。そして、MANACO をホスト側(石川地域側)とゲスト側(外部地域)に複数配置し、ゲストの観光情報閲覧・取得行動により発生するネットワークトラフィックを観測する。ホスト側観光資産に対するゲストの嗜好・興味など意味的特徴の

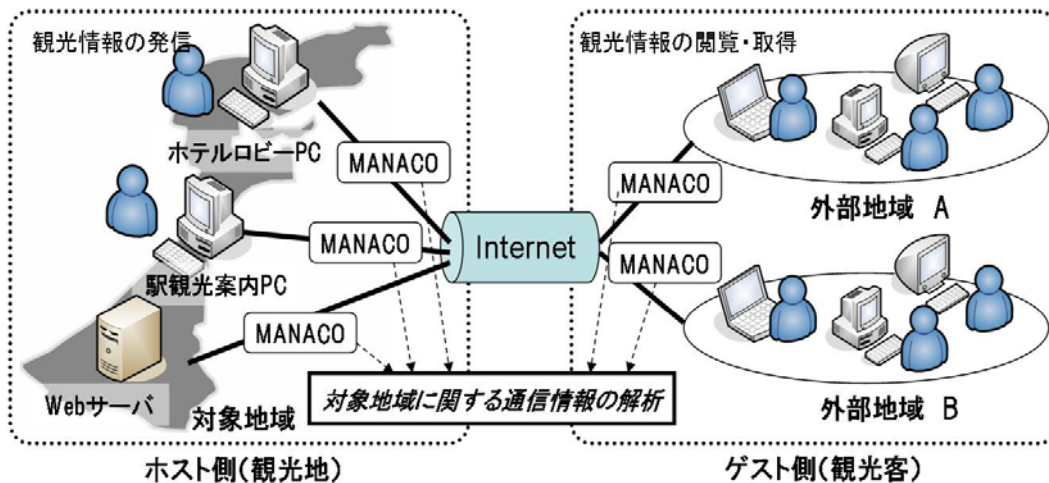


図2 地域観光情報観測の概要

分析・把握は、ゲスト側が適切な観光コンテンツを制作し観光ニーズに応えるという観点から重要であるが、その中でも本研究ではゲストが有する歴史的嗜好に着目し、ネットワークトラフィックからの史学的特徴の観測を行う。その理由は次の通りである。「地域内に現存する歴史観光資産の中で、観光客が興味を持っているのは一体どの時代のものか?」という問いは、歴史観光を推進する観光地域にとって極めて関心が高い。特に石川地域のように歴史資産の時代範囲が広くかつ資料数が多い観光地域では、ホスト側がゲストに対して的確に魅力ある歴史観光プランを提示することが、観光客誘致の増加に直結する。また、石川地域では季節・期間によって、歴史観光イベントや海産物などの名産品が多く提供されており、ゲストの観光目的もそれに伴い変化している。ゲストの歴史的嗜好を時系列に解析・把握することは、対象地域観光をより強くアピールする観点から重要である。

石川地域内の歴史観光資産を対象にしたネットワークトラフィックからの史学的意味の抽出を行う際には、抽出に際しては観測されたデータを予め定義された形式に基づいて整形し、HTMLデータが含まれている場合にはHTML部分の抽出をした後に、記述タグおよび日本語文字コードの判別を行う。整形された取得データから予め定義された石川歴史観光資産分類を用いて歴史資産名称単語を抽出するために、本研究では bi-gram マルコフモデル型の解析器である MeCab[9] を用いて日本語形態素解析を行った。そして、その結果に基づき、得られた歴史資産名称単語の取得データにおける出現頻度を算出する。石川歴史観光資産分類の定義に際しては、史学研究者の協力の下に作業を行い、定義された歴史観光資産の総数は金沢市を中心に約 500 件である。また、出現した歴史資産名称単語とその自体を関連付けるために、それぞれの歴史資産には出現年代に応じて「古代・中世・近世・近代・現代」の何れかの年代情報を定義した。表2に石川歴史観光資産分類を示す。

さらに、石川歴史観光資産分類に基づいた、ネットワークトラフィックの史学的特徴観測について、実際の観測結果例を図3に示す。この観測では、石川県の歴史観光資産を紹介した代表的な観光情報 Web サイト5箇所に対して、仮想アクセスソフトを利用した無作為 180 秒周期の情報取得

アクセスを実験用計算機から実行した。この情報取得アクセスにより発生したネットワークトラフィックを通信経路中から MANACO を利用して観測し、石川歴史観光資産分類に従って解析を行った。この観測では、「近世」に関する史学的特徴が多数観測されているが、これは対象とした観光情報サイトには「兼六園」「金沢城」など江戸時代の歴史資産に関する情報が非常に多く掲載されており、この Web コンテンツの史学的特徴が観測結果に反映されたものと思われる。

歴史資産名称	住所	時代	Key1
兼六園	金沢市兼六町 1-4	近世	史跡
ひがし茶屋街	金沢市東山	近世	史跡
長町武家屋敷跡	金沢市長町	近世	史跡
前田土佐守家	金沢市片町	近代	史跡
野村家	金沢市長町 1-3-32	近代	史跡
金沢城	金沢市丸の内 1-1	近世	史跡

表2 石川歴史観光資産分類 (一部抜粋)

まとめ

本研究では、インターネット上において発信・取得されている地域観光情報の観測および解析手法として、MANACO を用いた地域観光情報観測手法を提案した。地域観光情報の観測では、ホスト・ゲストという重要な関係性を意識する必要がある。観光客であるゲスト側が対象観光地域であるホスト側にどのような興味を抱いているかを定量的・客観的に明らかにすることが重要である。著者らが開発した MANACO は、ネットワーク通信路から直接的に通信が有する意味情報を抽出できるため、観光情報において最もゲスト側の興味を把握することが可能である。本論文では、観光情報に対する MANACO の有効性を示すために、石川地域内の歴史観光資産を対象にした史学的意味の観測を提案した。観測に際しては、石川歴史観光資産分類を定義し、それに基づいたネットワークトラフィックに

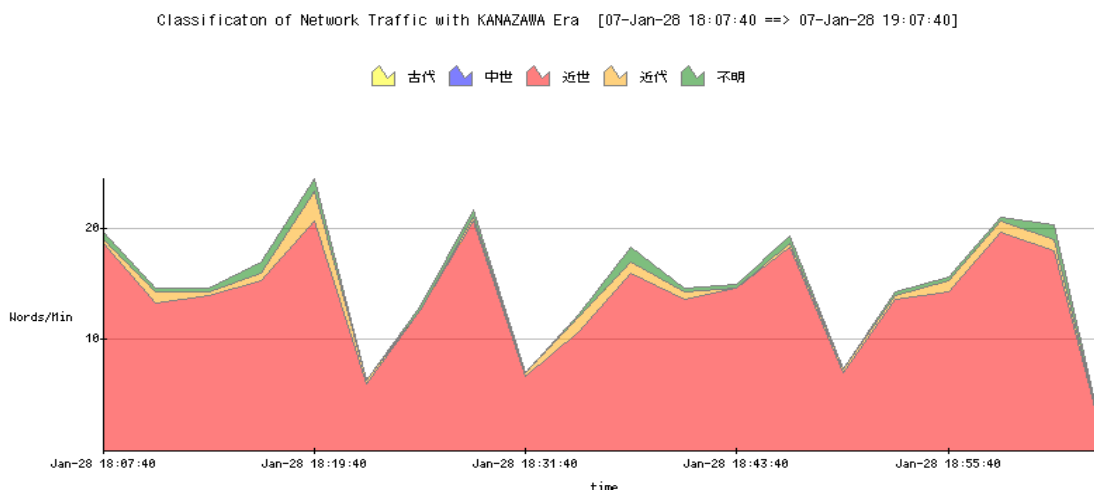


図3 石川歴史観光資産分類に基づいた史学的特徴観測

含まれる史学的特徴の把握を行った。その結果、特定の歴史観光 Web サイトの史学的内容がネットワーク上での観測結果においても反映されており、本提案の観光情報に対する有用性が明らかとなった。

謝辞

本研究は、文部科学省知的クラスター創生事業の一部として行われた。関係各位に謹んで感謝の意を表す。

参考文献

- [1] 国土交通省, 平成 18 年版 観光白書 (観光の状況に関する年次報告), 2006
- [2] 堀井, 林, 権, 吉田, メタデータ照合型医療情報通信監視システムの構築, 第 26 回医療情報学連合大会論文集, 2006
- [3] Smith, Valene (ed.), Hosts and Guests: The Anthropology of Tourism. Philadelphia: University of Pennsylvania Press., 1977
- [4] Google, www.google.jp
- [5] Page, L., Brin, S., Motwani, R. and Winograd, T., The PageRank Citation Ranking: Bringing Order to the Web, Technical report, Stanford University, 1998.
- [6] 中戸, 岩井原, ウェブ地域情報の自動要約のための特徴キーワード抽出, 電子情報通信学会データ工学ワークショップ (DEWS2005) 講演論文集, 2005
- [7] 加治屋, 山際, 蔵本, 正岡, ウェブサイト運営を通じた北の道のニーズ分析とブランドづくり～6 年間にわたる「北の道ナビ&北の道の駅」の運営の経験から～, 第 4 回 ITS シンポジウム論文集, 2005
- [8] RFC1157, A Simple Network Management Protocol (SNMP), <http://www.ietf.org/rfc/rfc1157.txt>.
- [9] MeCab: Yet Another Part-of-Speech and Morphological Analyzer., <http://mecab.sourceforge.jp/>.