



## Webページの信頼性の評価手法に関する研究

平林, 真実

---

(Degree)

博士 (工学)

(Date of Degree)

2004-03-31

(Date of Publication)

2010-02-23

(Resource Type)

doctoral thesis

(Report Number)

甲3075

(URL)

<https://hdl.handle.net/20.500.14094/D1003075>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



【 309 】

氏 名・(本 籍)	平林 真実	( 長野県 )
博士の専攻分野の名称	博士 (工学)	
学 位 記 番 号	博い第319号	
学位授与の 要 件	学位規則第4条第1項該当	
学位授与の 日 付	平成16年3月31日	

【 学位論文題目 】

Webページの信頼性の評価手法に関する研究

審 査 委 員

主 査	教 授	北村	新三
	教 授	上原	邦昭
	教 授	大月	一弘
	教 授	鏑木	誠

WWW が誰でも情報発信を行えるメディアとしての地位を確立するなか、個人の判断で情報発信が行えるという特徴ゆえに、正確でない情報、主観に頼った情報、内容が更新されていない情報などが数多く存在し、WWW コンテンツの信頼性が問題となっている。実際、信頼性の判断が難しい小さな企業のページが多数あり、それらを信用して取引したことによるトラブルなども発生している。さらに、WWW コンテンツの多くを占める個人の発信する情報は、どのような人が書いたのかが判断できなかつたり、内容が古いままであったりするなど、内容がどの程度信用できるのかわからないものが多い。

このため、WWW コンテンツを探す際に利用されている検索システムにおいても、信頼性を考慮して各ページにスコアを付ける方法が一般的になってきている。このようなシステムは広く利用され、コンテンツの判断において一定の成果を挙げている。しかし、スコアが高いページが必ずしも信用できるページであるとはいえず、信頼性判断には十分ではない。結局、内容の信頼性の判断は、内容を読んだ個々による判断に任されているのが現状である。

本研究では、Web ページの信頼性を評価する方法について検討する。

一般に WWW コンテンツの信頼性を評価する際には、次のような問題があると考えられる。

1 信頼性の基準を一意に決定することが難しい

信用できるページ作者や考え方などの信用の基準は様々な観点から設定可能であり、評価する人により異なる可能性がある。

2 内容からのみからコンテンツの信頼性を判断することが難しい

内容からはテーマや伝えたいことなどは判断できるが、その内容について詳しく知っている場合やページの作者を良く知っている場合を除いて、内容が信用できるかを評価することは難しい。したがって、キーワード検索や構文解析などは、「目的とする内容を記述してあるページ」を探すことには有効であるが、信頼性を評価することにはあまり適していない。

これらの問題に対し、本研究では以下のような考えにより信頼性の評価を行う。

まず、信頼性に対しどのような基準が可能であるかについて考える。我々が通常利用している信頼性の基準について考えると、誰が作成したページであるかなどの作成者に関する判断に基づいて信頼性を評価することはよくあることである。例えば、官公庁や有名な企業あるいは良く知っている人によって書かれた内容ならば一定の信用を置くことができるし、知らない人の書いた内容は一概には信用できないといった判断を行っている。このような信頼性の判断は WWW からではなく評価する人の社会的な知識に基づいて行われ、その結果がページの判断に反映されると考えることができる。本手法では、このような我々が通常行っている信用についての評価、つまり、WWW 以外で決定される外部評価をペー

ジ信頼性の基準として用いる。

次に、2 の問題に関しては、内容に依存しない評価方法が必要であると考えられる。このような方法の一つとして、内容を解析せずにページを評価する方法であるソーシャルフィルタリング的な手法が目玉されている。代表的な手法として、WWW 構造であるリンク関係を解析する方法がある。現在、最も利用されている検索エンジンである Google のページ評価の中心となる PageRank 法では、リンク解析において「多くの良いページからリンクされているページは良いページである」という考えに則り、マルコフ過程を用いてリンクをたどる際のページの滞在確率を求め、各ページの評価として点数化する方法を採用している。この方法はページの人気度を評価することで信頼性の向上に貢献しているといえる。ただし、どの程度信用できるかという信頼性そのものを評価したものではないため、点数が高いページの方が必ずしも信頼が高いページでないという問題がある。

本研究で提案するページ評価手法では、WWW におけるグラフ構造を一種のネットワークフローの問題として捉え、初期値としてページに対する信頼度を指定し、ページ間のリンク構造の解析を行うことで、自動的に各ページに対し信用に基づいた信頼性の評価を得るというアプローチを採用する。

本提案手法の特徴として、以下の2点がある。

- ・信頼性の基準となるものに WWW 以外の外部により決定される「社会的信用」の評価を用いて、一部のページに対し評価を与える。すなわち、既知のページの社会的な信頼性を外部評価として利用する。

- ・リンク構造解析を利用して、それらのページの評価からすべてのページの評価を求める。すなわち、リンクによる信頼性が伝播するという考えに基づきページ評価を行う。

本手法における外部評価として社会的知識に基づく信頼である「社会的信用」という概念を導入することで、誰が信用できるかという判断において、客観的評価の中にある程度の主観性を含んだ信頼性の評価を行うことができる。主観を含む評価を許すことにより、統一的な評価とともに多様な側面を持つ広義の信頼を評価することも可能にしている。

リンク構造解析においては、リンクの意味をどのように捉えて信頼性の判断に利用するかがページ評価において重要となる。多くの場合、我々がリンクを作成するのはリンク先ページの内容を評価した上で見る価値がある、あるいは見てほしいと判断した場合である。このようなリンクは一種の推薦を意味するものと考えられる。リンクによる推薦を信頼という観点から見ると、リンクを作成した人はリンク先ページに対し推薦した責任を持つ、すなわち、推薦することで信頼を付与していると捉えることができる。これは、リンクを介してページへ信用が伝播していると考えられる。本手法では、このような考えに基づきリンク構造の解析を行うことで各ページの評価を行う。

本論文の構成は以下の通りである。

第2章では WWW 以外の外部情報としてページ作成者の社会的信用という概念を導入し、社会的信用に基づいた Web ページの評価を一般化モデルとしてグラフ理論を用いて数学的に定義し各ページの評価を導出するアルゴリズムを提案する。

本モデルでは、

- 1 いくつかのページに対しては信頼性が評価されている
- 2 リンク関係をリンク元ページからリンク先ページへの推薦と解釈する
- 3 信頼度の高いページから推薦されたページの信頼度は高くなる。すなわち、リンクを介して信頼度が伝播するという仮定の妥当性をしめす。

さらに、ページ作成者の URL と信用度を指定することと、グラフ理論を用いてリンク構造解析を行うことで、各ページの信頼度を計算する手法および WWW へ適用するためのアルゴリズムを提示する。本アルゴリズムはダイクストラ法を改良することにより、妥当な計算量によりすべてのページの評価を行うことができる。また、提案方法の汎用性に関する議論を行い、一般化モデルで使用するパラメータを調整することで、信頼性を表すいくつかの代表的な基準に適用できることを示す。

第3章においては、大規模な Web サイトにおけるサイト構造に注目し、Web サイト内のページの信頼性を評価するための手法について述べる。ここでは、大学や研究所などの組織が持つサイトにおいて多くの作成者が、各々のページを作成する環境を想定し、信頼性の基準として公式度という概念を導入する。組織においてはページを作成する人や部署に応じて、情報を発信する際の責任や裁量が暗黙あるいは明確に規定されており、誰の責任でページを作成したかによってページの公式度が異なっている。また同一作成者によるページでも、公式ページや趣味のページなどで明らかに公式度が異なるページも存在している。

3章では、第2章で提案した一般化モデルにおいて、

- ・組織構造を利用して各ページの初期値を決定する
- ・リンクに属性を付与することで、リンクの意味を明示させる

ことにより、サイト内の各ページの公式度を導出する方法を提案する。

さらに、実在のサイトを対象とした評価実験を行い、提案手法の妥当性と有効性を明確にしている。また、提案手法を利用することにより、現在の企業のような公式情報しか存在しないサイトにおいても自由度の高い情報発信が可能になることを示した。

第4章では、社会的信用に負の信用という概念を導入し、正と負の信用から WWW 全体における良いページと悪いページを判定する手法の提案する。ここでは、良いページと悪いページの持つ性格によりリンクの性質に違いがあることを利用し、リンクによるページ関係をモデル化し、良いページ/悪いページ/悪い可能性のあるページを判定する手法を導出

した。

第2章で示した方法に対し、

- 1 評価者が良いページ、悪いページと判断されるページ群を指定する。
- 2 すべてのページに対して推移的閉包を用いてリンクを解析することで、良いページ群/悪いページ群との到達可能性を調べる
- 3 得られた到達可能性の関係によって、各ページの良い/悪い/gray(悪い可能性)の判定を行う

ことにより、各ページの信用を判定している。

さらに、実在のサイトに対して調査を行い本判定基準の妥当性であることと、ポータルサイト、検索サイト、ランキングサイトなどでは様々なページに網羅的にリンクしているために、本手法においては取り除いて解析するべきであることを明らかにした。また、WWW に適用する場合の有効性と問題点を示すことで、WWW におけるリンク関係についての考察を行う。

第5章では、本研究におけるページ信頼性の評価手法について総括を行う。

氏名	平 林 真 実		
論文 題目	Web ページの信頼性の評価手法に関する研究		
審査委員	区 分	職 名	氏 名
	主 査	教 授	北 村 新 三
	副 査	教 授	上 原 邦 昭
	副 査	教 授	大 月 一 弘
	副 査	教 授	鏑 木 誠

甲

## 要 旨

## 概要

WWW が誰でも情報発信を行えるメディアとしての地位を確立するなか、個人の判断で情報発信が行えるという特徴ゆえに、正確でない情報、主観に頼った情報、内容が更新されていない情報などが数多く存在し、Web コンテンツの信頼性が問題となっている。このため、インターネットでコンテンツを探す際に利用されている検索システムにおいても、信頼性を意識して各ページにスコアを付ける方法が一般的になってきている。このようなシステムは一定の成果を挙げているが、まだ信頼性の判断には十分であるとは言えず、結局、内容の信頼性の判断は、内容を読んだ個々による判断に任されているのが現状である。

本論文では、Webページの信頼性を評価する方法を開発することを目的とする。

第1章では、まず、本研究の目的、動機、ページの信頼性の考え方、ページ評価に対する従来研究の紹介が述べられ、これに対して本研究によって期待される効果について説明されている。

特に、Webページの信頼性に関する評価尺度については、従来、あまり検討が行われていない部分であるが、本研究では、以下のように基本概念を定義している。本来、情報に対する信頼度の評価尺度を一意に定めることは困難であり、個々の主観や視点によっても尺度は異なる。そこで、ここでは、何らかの基準をもととする評価に対して、各ページの相対的評価を行うことを考える。何らかの基準を表す概念として、社会的信用という概念を導入する。例えば、官公庁や有名な企業あるいは良く知っている人によって書かれた内容ならば、一定の信用を置くことができるという判断基準は、一般に多くの人々にも受け入れられている基準である。このような信頼はWWW上のコンテンツからではなく、評価する人の社会的な知識に基づいて決定されるものである。

第2章では、社会的信用に基づいたWebページの評価を一般化モデルを数学的に定義し、各ページの評価を導出するアルゴリズムを提案している。ここでは、まず、

- (1) いくつかのページに対しては信頼性を評価されている。
- (2) ハイパーリンクをリンクもとのページのリンク先ページに対する推薦と考える。
- (3) 信頼度の高いページから推薦されたページの信頼度は高くなる。即ち、リンクにより信頼度が伝播する。

という仮定の妥当性を示している。更に、ページ作成者のURLとその信用度によりページ作成者に対する信用度を指定し、グラフ理論を用いてページとリンク間の構造解析を行うことで各ページの信頼性を計算する方法、ならびにそのアルゴリズムを提示している。更に、提案方法の汎用性に関して議論を行い、一般化モデルで使用するパラメータを調整することで、信頼性を表すいくつかの代表的な基準に適用できることを定性的に示している。

第3章においては、大規模なWebサイトにおけるサイト構造に注目し、Webサイト内のページの信頼性を評価するための手法について述べている。ここでは、大学や研究所など組織を持つサイトにおいて多くの人間が各自ページを作成することを想定し、また、信頼性の基準として公式度という概念を導入している。組織において、ページを作成する部署や人に応じて情報を出す際の責任や裁量が暗黙あるいは明確に規定されており、誰の責任でページを作成したかによって、ページに対する公式度が異なる。また、同一人物が作成したページにおいても、公式ページ、趣味のページなど明らかに公式度が異なるページが存在する。本章では、2章で提案した一般化モデルに対して、

氏名	平 林 真 実
----	---------

(1) 組織の組織構造を利用して各ページの初期値を決める。

(2) リンクに属性を付与することで、リンクの意味を明示させる。

ことにより、サイト内の各ページの公式度を導出する方法を提案している。更に、実在するサイトを対象とした評価実験を行い、提案方式の妥当性・有効性を明確にした。また、提案方式を利用すれば、現在の企業のような公式情報しか存在しないサイトにおいても自由度のある情報発信が可能になることを示した。

第4章では、社会的信用に負の信用という概念を導入し、正と負の信用からWWW全体における良いページと悪いページを判定する手法の提案を行っている。ここでは、良いページと悪いページの持つ性格によりリンクの性質に違いがあることを利用し、リンクによるページ関係をモデル化しているところに特徴がある。2章で提案した方式を利用して、

- (1) 評価者が良いページ、悪いページと判断されるページ群を指定する。
- (2) すべてのページに対して、推移的閉包を用いてリンクを解析し、良いページ群、悪いページ群からの到達可能性を調べる。
- (3) 正負のページからの到達可能性の関係によって、各ページの良い/悪い/gray(悪い可能性)の判定を行う。

ことにより、各ページの信用を判定している。さらに、実在のサイトに対して調査を行い本判断基準が妥当であること、ポータルサイト、検索サイト、ランキングサイトなどは、様々な種類のページを網羅的にリンクしているため、本手法による解析時にはそのサイトを削除するべきであることを明らかにした。

第5章は、本研究におけるページ信頼性の評価手法について結果をまとめたものである。

以上のように、本研究は近年話題となっているWWWページの信頼性の問題に関して、信頼度を数値的に評価する方法を研究したものであり、WWWの利用を効果的かつ広範囲に広げていく上での重要な知見を得たものとして価値ある集積であると認める。

よって、学位申請者の平林真実 は、博士(工学)の学位を得る資格があるものと認める。