重要ブロガー判定手法を利用した意見分析サイトの構築

桑原 雄† 稲垣 陽一† 中島 伸介‡ 張 建偉‡

†株式会社きざしカンパニー 〒103-0015 東京都中央区日本橋箱崎町 24-1
‡京都産業大学 コンピュータ理工学部 〒603-8555 京都府京都市北区上賀茂本山
E-mail: †{yk,inagaki}@kizasi.jp, ‡nakajima@cse.kyoto-su.ac.jp, ‡zjw@cc.kyoto-su.ac.jp

あらまし ブログは特定のブロガーが記述するコンテンツであり，過去の投稿履歴を解析することで，そのブロガーの特性を分析することが可能である．本研究では，このブログの特性に着目し，重要なブロガーの発見および分析手法を応用して構築した，意見分析サイトの紹介を行う．

キーワード ブログ解析，Web マイニング

Introduction of Opinion Analysis Website Based Upon Influential Blogger Detection Method

Yu KUWABARA† Yoichi INAGAKI† Shinsuke NAKAJIMA‡ and Jianwei ZHANG‡

†kizasi Company,Inc 24-1 Hakozaki-cho, Nihonbashi, Chuo-ku, Tokyo, 103-0015 Japan
‡Kyoto Sangyo University Motoyama, Kamigamo, Kita-ku, Kyoto-shi, Kyoto, 603-8555 Japan
E-mail: †{yk,inagaki}@kizasi.jp, ‡nakajima@cse.kyoto-su.ac.jp, ‡zjw@cc.kyoto-su.ac.jp

Abstract Blogs are often very personal in nature, reflecting the very character of their blogger. Thus, by analyzing a blog's past entries, it is possible to better understand a blogger's personality. In this research, we focus on blogger's character, and introduce our opinion analysis site which employs an influential blogger detection method.

Keyword Blog Analysis, Web Mining

1. はじめに

近年，ブログに代表される CGM と呼ばれるコンテンツが大量に配信されるようになっている．ブログの特性として，“あるブログサイトの全てのエントリは，基本的に 1 人のブロガーが記述したものである”というものがある．よって，過去に投稿されたエントリの履歴を解析することで，そのブロガーの特性を把握することが可能であると考えられる．逆に，ブロガーの特性を把握することができれば，共通した特性を持つブロガーの集合を作成することで，ブロガー全体ではなく，その領域に関してより深い知識／興味を持った（重要な）ブロガーを対象とした分析が可能になると考えられる．

我々は，特定トピックにおける重要なブロガーを発見し，その集合を分析することで，この特定トピックに関する有益な知見が得られると考え，対象トピックとして競馬を扱った Web システムを開発した．また，本システムを利用した検証実験を行ったので併せて紹介する．

2. 重要ブロガーの発見

本稿では対象トピックとして“競馬”を扱う．これは予想を書くブロガーが多いことと，その予想結果の正誤の判定が正確に行えることから，提案手法の検証を行うために適切と判断したからである．

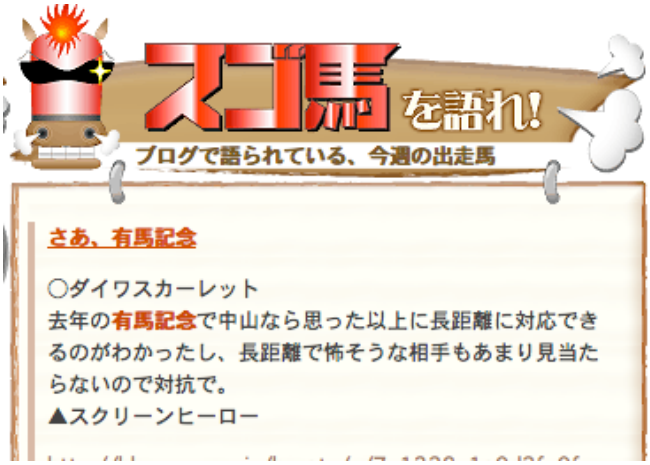

図 1 構築した Web サイトでのエントリの表示

ここでは競馬というトピックにおける，ブロガーの

重要度の算出方法について述べる．まず，各ブロガーがレースを予想した回数と予想が的中した回数をブログから取得し，それらに基づいて重要度を算出した．

競馬の予想をするブログエントリでは，馬名と合わせて「◎」や「△」などの記号を表記することで自分の予想を示すことが多い．図 1 に，競馬の予想をしているエントリの例として，今回構築したサイトにて表示されているものを示す．

レース開催前に投稿されたエントリに対して，レース名，出走馬名，記号を用いたパターンマッチングを行い，予想回数と的中回数を自動的に取得した．

ここで，あるブロガーの対象トピックにおける重要度（score）を以下のように定義した．

$$score = \left(\frac{H}{F}\right) \times H^{0.75}$$

なお，F は予想回数，H は的中回数である．なお，$H^{0.75}$ により，予想回数が少ないブロガーのスコアが高くなってしまわないよう，的中回数を用いて重み付けを行っている．この score によりブロガーのランク付けを行い，閾値以上のブロガーを重要ブロガーと判定した．

3. 評価実験

3.1. 使用したデータ

2005 年 4 月～2008 年 11 月のレース結果（全 1043 レース）を準備し，ランダムに選択した 30 レースを予想対象とした．残りの 1013 レースを重要ブロガー発見のためのデータとして使用した．

3.2. 手順

（1）1013 レースのデータを用いて，各ブロガーの score を算出する．今回は，score の上位 300 人を重要ブロガーとした．

（2）重要ブロガー300 人のエントリを用いて，レース名，出走馬名，記号を用いたパターンマッチングを行い，レースの予想を作成する（予想 A）．

（3）比較対象として，重要ブロガー300 人以外の全ブロガーのエントリを用いて同様の予想を作成し（予想 B），結果を比較する．

ここで，本命の馬を指す場合は「◎」や「○」，要注意の馬を指す場合は「△」や「▲」を用いることが多いため，それぞれ異なるパターンマッチングを行い，本命の馬 2 頭と要注意の馬 2 頭の計 4 頭を予想した．レース結果の上位 2 頭がその 4 頭に含まれている場合，予想が的中したとみなした．

3.3. 結果

予想対象のランダム選択を繰り返す事で，合計 10 回の試行を行った．図 2 にその結果を示す．

試行回数 10 回の内，9 回の試行で重要ブロガーの予想の的中回数が多かった．また，的中率では 8.7%の差がみられた．よって，重要なブロガーを集めて分析することで，ブログ全体を分析するよりもより有益な知見が得ることができたと言える．

	1	2	3	4	5	6	7	8	9	10	計	的中率
A	11	10	7	6	10	7	8	11	9	8	87	29.0%
B	5	8	5	8	9	6	6	5	3	6	61	20.3%

図 2　30 レースを予想した場合の的中回数（10 回試行）

4. サイトの構築

今回行った実験を元に，競馬の予想を重要ブロガーのブログエントリから自動的に作成し，コンテンツとして表示するサイトの構築を行った．

図 3　サイトでの予想の表示

図 3 に，重要ブロガーのブログエントリから作成した，本命の馬（スゴ馬）と要注意の馬（ダーク馬）のランキングを示す．評価実験ではそれぞれ 2 頭を選んだが，サイトではそれぞれ 5 頭ずつを選び，ランキン

グ形式で表示している．スゴ馬 5 頭は記号の数，ダーク馬 5 頭は「◎」や「○」に対する「△」や「▲」の比率を元に算出している．

5. 今後の課題

- 重要ブロガー発見手法の改善

既存研究として，話題に対する熟知度を測定し，より信頼性の高い情報を提示する研究[1]が行われている．ここで提案されている技術を応用することで，重要ブロガー発見手法の改善に向けた検討を行う事を予定している．

- 他のトピックへの応用

競馬だけでなく，その他のトピックでも同様の結果が得られるか，さらに実験を進める必要があるといえる．

6. 終わりに

本稿では，特定領域において重要なブロガーを発見し，その集合を分析することで，ブロガー全体を分析対象とするよりも有益な知見が得られる，という仮説を検証し，その結果を応用して構築した Web サイトを紹介した．今後は，よりよい重要なブロガーの発見手法を検討するとともに，他のトピックでの実験を進めていく予定である．

謝　辞

本研究の一部は，NICT 委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」によるものです．ここに記して謝意を表します．

参　考　文　献

[1] 中島伸介，稲垣陽一，草野奉章，“高信頼性情報の提示を目指した熟知度に基づくブログランキング方式の提案，”日本データベース学会論文誌，Vol.7，no.1，pp.257-262，Jun. 2008.