

テレビ番組に基づく Web コンテンツの信頼性判断支援システムの提案

佐伯 隆太[†] 宮森 恒[†]

[†] 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: †{g0846589,miya}@cse.kyoto-su.ac.jp

あらまし 近年、情報を収集する手段として、Web コンテンツが広く利用されている。Web コンテンツはその量や多様性といった面で非常に膨大であるが、有用な質の高い情報と、嘘やデマなどの質の低い情報が混在しているという問題がある。一方、テレビ番組は、従来からプロの番組製作者によって作成されており、公共性や視聴率を確保する必要があるため、常に一定の品質が維持されていると考えられる。本稿では、テレビ番組の高い品質に着目し、これと Web コンテンツを比較することで、Web コンテンツの信頼性判断を効率よく支援するシステムを提案する。特に、本稿では、ある対象の時間的変化を記述した動向情報に絞り、これをテレビ番組中の関連した図表画像と比較するシステムを実装した。実験の結果、網羅度は低いものの、MAP 値としては 0.9 以上の高い値を確認した。

キーワード Web コンテンツ, 情報信頼性, 動向情報, テレビ番組, 図表画像

Proposal of a Credibility Decision Support System of Web Contents based on TV Programs

Ryuta SAEKI[†] and Hisashi MIYAMORI[†]

[†] Faculty of Engineering, Kyotosangyo University
Kamigamo-Motoyama, Kita, Kyoto, 603-8555 Japan

E-mail: †{g0846589,miya}@cse.kyoto-su.ac.jp

Abstract In recent years, Web information has been widely used as a means to collect information. Although Web information is very huge in quantity and diversity, its quality is a mixture of good and poor. Some information is truly useful to our daily lives, and some of it has no basis in truth. In contrast, television programs have been created by professional programmers, and they offer excellent quality in general due to an obligation to serve public interest and to maintain audience rate. In this paper, a credibility decision support system of Web contents is proposed by making it easier to compare Web contents with related TV programs, focusing on the high quality of content provided by TV programs. Here, in particular, the queries are focused on the trend information which describes temporal situations or changes about certain objects or events, and the system was developed comparing the query with related chart pictures in TV programs. Experiments showed that the precision indicates fairly high with 0.9 or more in MAP value, whereas the degree of coverage was quite low.

Key words Web contents, Information reliability, Trend information, TV program, Chart picture

1. はじめに

近年、情報を収集する手段として、Web コンテンツが広く利用されており、私たちの生活を手助けしている。Web コンテンツは即時性に優れ、量、多様性が非常に膨大といった性質をもつが、その一方で、有用な質の高い情報と、嘘やデマなどの質の低い情報が混在しているという問題がある。例えば、Web 上には、ニュース、ブログ、企業のホームページ、2ちゃんねるな

ど、様々な発信者からのページがある。社会的評価の高い企業のホームページや大手新聞社が配信するニュースの情報は比較的信頼度が高いと考えられるが、個人サイトやブログ、2ちゃんねる上の情報は信頼度のばらつきが大きく、その判断も容易にはできない。この問題を解決するために、これまでに、特定の Web コンテンツについて、その発信者や社会的評価、主要・対立・対比文といった異なる観点からの関連情報を提示するシステムや、対象となる Web コンテンツを何らかのデータ対で

表現し、そのデータ対と他のデータ対との関係を分析することにより該当 Web コンテンツの信頼性判断を支援する手法などが提案されている。

本稿では、テレビ番組を一定の品質が確保された情報と考え、これに基づき Web コンテンツの信頼性判断を支援するシステムを提案する。具体的には、Web 情報における動向情報を対象に、それと関連するテレビ番組の図表画像を提示することで、その動向情報の信頼性判断を効率よく支援することを目指す。動向情報とは、一般に、商品の売り上げ推移や内閣支持率の変化等、ある事柄や数値についての時間的変化を表現するデータあるいは言語情報を指すが、本稿では特に後者の言語情報で表現されたものを対象とする。また、図表画像とは、その中に折れ線グラフや円グラフ等の図表が主として出現している画像フレームのことである。テレビ番組は、従来からプロの番組制作者によって作成されており、公共性や視聴率を確保する必要があるため、常に一定の品質が維持されていると考えられる。よって、提案システムにより、Web コンテンツにおける動向情報の信頼性判断の効率化が期待される。

2. 関連研究

Web コンテンツの信頼性判断支援の研究は数多く行われている。

その一つに、Web コンテンツの信頼性を判断する基準として、「情報内容、情報発信者、情報外観、社会的評価」といった4つの観点から関連情報をユーザに提示することで、Web コンテンツの信頼性判断を支援する研究 [1] がある。情報内容については、Web ページの本文に書かれている内容に着目しており、情報発信者については、発信者の身元に着目した所属の分類やその分野での専門性の有無に着目している。情報外観については、情報ソースやデザイン、連絡先の明記などの Web ページの外観に着目しており、社会的評価については、他の利用者がその情報についてどのような見方をしているかに着目している。

上記の各観点からの情報の自動抽出に関連する研究も行われている。例えば、情報内容については、文中の同義、対比、対立関係に着目することで、文書の主要・対比・対立分を抽出する研究 [2]、情報発信者については、Web ページの主要部と情報発信者名の出現位置の関係に着目することで発信者を同定する研究 [3]、情報外観については、出現位置や共起文字列等のパターンをページタイプ分類と融合的に利用して抽出する研究 [?]、社会的評価については、評価表現中の各文節を隠れ変数に対応付けた条件付き確率場を用いた評価極性分類手法 [5] が挙げられる。

またデータ対間の関係分析に着目した Web コンテンツの信頼性評価の研究が挙げられる [6]。ここでは、評価対象の Web コンテンツとそれに関連する Web コンテンツのデータ対で表現し、その support 関係の強さで対象コンテンツの信頼性を評価するというモデルを導入している。

この他にも、様々な関連研究が行われているが、動向情報の信頼性を判断する基準としてテレビ番組を利用するという点で本研究は従来研究と異なる。

3. 提案システム

3.1 システム構成

本システムの構成図を図 1 に示す。本システムは、インデッ

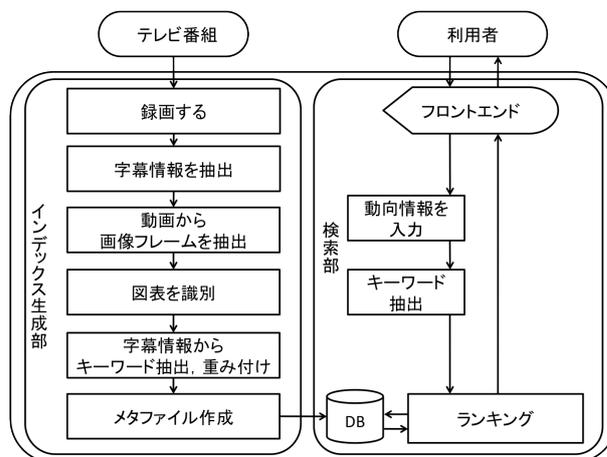


図 1 システム構成図

クス生成部と検索部で構成される。

インデックス生成部は、図表画像、キーワードを抽出し、メタファイルを生成する。以下の手順で処理を行う。

- (1) テレビ番組を録画して動画、字幕情報（クローズドキャプション）を抽出する
- (2) 動画から一定間隔おきに画像フレームを抽出する
- (3) 図表画像識別器を用いて、抽出した画像フレームに図表が含まれるかどうか判定する（図表が含まれると判定されたフレームを図表画像と呼ぶ）
- (4) 字幕情報から、図表画像の前後 N 秒に出現するキーワードを抽出し、重み付けする
- (5) 抽出した図表画像、重み付けしたキーワードをインデックス化し、DB に格納する

検索部は、利用者が与えたクエリを受け取り、DB 内の図表画像をランキングし、結果を出力する。大まかな処理手順は以下の通り。

- (1) 利用者が入力した動向情報クエリからキーワードを抽出する。
- (2) クエリから抽出したキーワードと、字幕情報から抽出したキーワードの合致度によって、図表画像をランキングする
- (3) ランキング結果を出力する

本節の以降では、システムにおける処理手順の詳細を述べる。まず、インデックス生成部の手順 (3) の詳細を 3.2 で述べる。次に、手順 (4) の詳細を 3.3 で説明する。最後に、検索部の手順 (1)、(2) の詳細を 3.4 で述べる。

3.2 図表画像識別器の構築

本稿では、テレビ番組の画像フレーム中に図表が含まれるかどうかを判定する識別器を構築するために機械学習の枠組みを利用する。

3.2.1 正例, 負例画像の収集

録画したテレビ番組, および, Web から人手で図表画像を収集し, 正例とする. 図表が使用されている番組のほとんどがニュースであったため, 今回は, ニュース番組を中心に作業を進めた. 図表の種類は, 折れ線グラフ, 棒グラフ, 円グラフを対象 (図2) とし, それぞれについて正例を収集した. 同様に, 負例もテレビ番組, および, Web から人手で収集した. 以上により, 表1のような正例, 負例を収集した. 本稿で利用したテレビ番組を表2に示す.



図2 図表画像の例

表1 正例, 負例の件数

	正例数	負例数
円グラフ	200件 (テレビ:74件, web:126件)	1300件 (テレビ)
棒グラフ	130件 (テレビ:74件, web:56件)	1300件 (テレビ)
折れ線グラフ	180件 (テレビ:157件, web:23件)	1300件 (テレビ)

表2 利用したテレビ番組

番組名	テレビ局	放送時間	録画時間の合計	収集した正例数
NHK ニュース おはよう日本	NHK 総合	07:00~07:45(毎日)	147時間	51件
Nスタ	毎日テレビ	17:00~18:15(平日)	70時間	40件
VOICE	毎日テレビ	18:15~19:00(平日)	105時間	38件
NHK ニュース7	NHK 総合	19:00~19:30(毎日)	98時間	81件
ニュースウォッチ9	NHK 総合	21:00~22:00(平日)	140時間	62件
ワールドビジネスサテライト	テレビ大阪	23:00~23:58(平日)	135時間	33件

3.2.2 図表画像の検出に有効な特徴量と学習

近年, 物体検出に関する研究が盛んに行われている. 物体検出の代表的な対象は顔であり, 近年では物体検出の実用化が進んでいる [7] [8] [9]. 物体検出には, 対象の物体に共通な要素を最小限抽出する必要がある, その際, 抽出する要素が特徴量である. 本稿では, 図表画像の検出方法として, HOG+AdaBoost を使用した [10] [12]. HOG 特徴量は, エッジ情報に着目した特徴量であることから, 図表画像の検出に有効であると考えられる.

HOG+AdaBoost では, HOG 特徴量と AdaBoost を用いて図表画像を検出する. HOG 特徴量は, 一つの局所領域にお

けるエッジ情報に着目した特徴量であり, 輝度勾配の方向についてヒストグラムをとったものである. AdaBoost は, 機械学習における学習アルゴリズムの一つであり, ひとつひとつはあまり判別能力の高くない弱識別器の重み付け多数決をとることで強い識別器を作る手法である. HOG 特徴量と AdaBoost を組み合わせることで, 図表画像を正確に判別することを目指す. HOG+AdaBoost での図表検出の流れを図3に, HOG 特徴量抽出の流れを図4に示す. なお, 画像からグラフに該当する部分を切り取り, 128x128のサイズに正規化した画像を学習データとした.

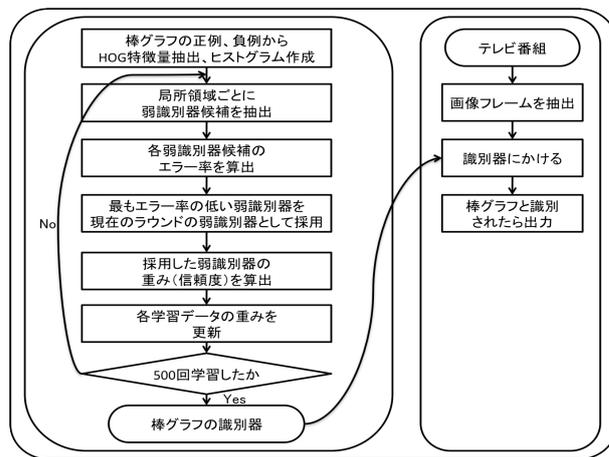


図3 HOG+AdaBoost での図表画像検出の流れ (棒グラフの場合)

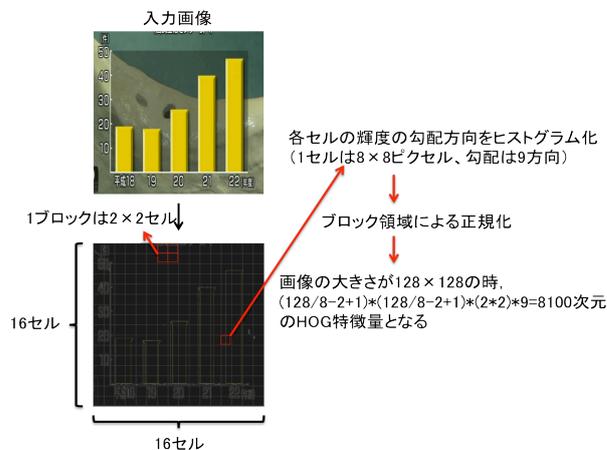


図4 棒グラフの正例, 負例から HOG 特徴量抽出, ヒストグラム作成

3.3 字幕情報からのキーワード抽出と重み付け

3.3.1 キーワード抽出

図表画像と判定されたフレーム番号を中心に, その前後15秒の字幕情報を図5のように抽出する. ここで, フレーム番号と時刻は, テレビ番組のフレームレートが29.97 (フレーム/秒)であることを踏まえて対応付ける. 以上のように抽出した字幕情報を MeCab [13] で形態素解析し, 単名詞, 複合名詞を抜き出し, 図表画像の関連キーワードとする. なお, mecab でキーワードを抜き出す際, 動詞, 形容詞等の自立語をキーワードとして用いると, 不要なキーワードが抽出されてしまうといった

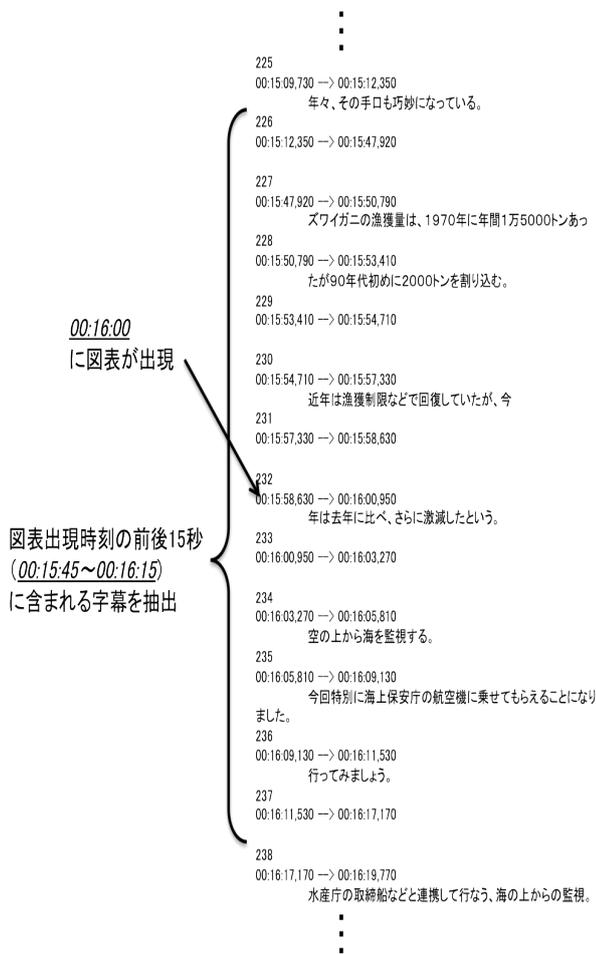


図5 文字列の抽出

問題が生じることを予備実験で確認した。よって、ここでは、単名詞、複合名詞のみを関連キーワードとして抜き出すこととした。

3.3.2 重み付け

キーワードの重み付けには、*tf-idf*, *FLR* 法を用いる。

tf-idf は、文書検索で用いられる語の重要度算出手法であり、*tf* (*Term Frequency*) と *idf* (*Inverse Document Frequency*) の積によって計算される。文書 *d* におけるキーワード *k* の *tf* は以下の式によって算出する。

$$tf(d, k) = \frac{n_k}{W_d} \quad (1)$$

ここで、 n_k はキーワード *k* の出現回数であり、 W_d は文書 *d* における全キーワード数である。また、キーワード *k* の *idf* は、

$$idf(k) = \log \frac{N}{w_k} + 1 \quad (2)$$

で表される。ただし、*N* は総文書数、 w_k はキーワード *k* を含む文書数である。以上より、文書 *d* におけるキーワード *k* の重要度である *tf-idf*(*d, k*) は、以下の式によって算出される。

$$tf-idf(d, k) = tf_{d,k} \times idf_k \quad (3)$$

FLR 法は、中川らが提案した複合名詞に対する専門用語らしさの算出手法 [11] である。まず、複合名詞 *CN* のスコア

LR(*CN*) を以下の式によって求める。

$$LR(CN) = \left(\prod_{i=1}^L (FL(N_i) + 1)(FR(N_i) + 1) \right)^{\frac{1}{2L}} \quad (4)$$

L は複合名詞 *CN* における単名詞数、*FL*(*N_i*) は単名詞 *N_i* の左方に別の単名詞が接続する頻度、*FR*(*N_i*) は単名詞 *N_i* の右方に別の単名詞が接続する頻度である。また、*LR*(*CN*) を利用し、*FLR* 法は以下の式で算出される。

$$FLR(CN) = f(CN) \times LR(CN) \quad (5)$$

f(*CN*) は複合名詞 *CN* が単独で出現した頻度である。

tf-idf, *FLR* 法でそれぞれ重み付けされたキーワードを、上位 5 件、または、上位 10 件のみそれぞれ用いる場合と比較した結果、全キーワードを用いる場合の方が良好な結果が得られたため、ここでは、*tf-idf*, *FLR* 法それぞれの全キーワードを図表画像の関連キーワードとして DB に登録する。

3.4 検索

本システムでは、利用者から与えられるクエリは、日本語の自然文を想定している。利用者は、Web 上の調べたい動向情報をコピー・アンド・ペースト等により、クエリ入力用のテキストボックスに入力する。入力例を図 6 に示す。

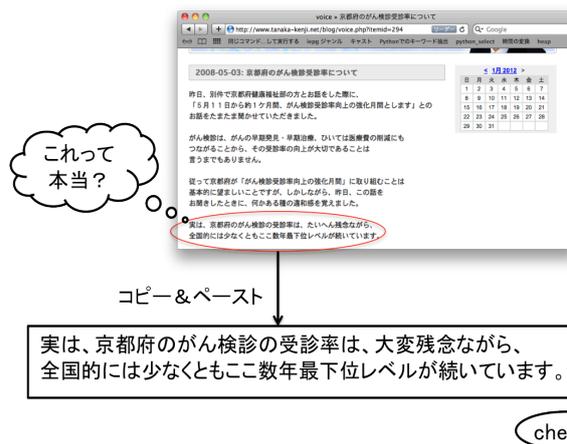


図6 入力例

入力されたクエリから単名詞、複合名詞を抽出し、得られた名詞集合からクエリベクトルを生成する。mecab でキーワードを抜き出す際、3.3.1 と同様に、動詞、形容詞等の自立語をキーワードとして用いると、不要なキーワードが抽出されてしまうといった問題が生じることを予備実験で確認した。よって、ここでは、単名詞、複合名詞のみをキーワードとして抜き出す。検索結果は、ベクトル空間モデルによるコサイン類似度でランキングする。クエリベクトル V_q と図表画像に対応する文書ベクトル V_d のコサイン類似度は以下の式で表される。

$$\cos(V_q, V_d) = \frac{V_q \cdot V_d}{\|V_q\| \|V_d\|} \quad (6)$$

ここでは、クエリベクトルの重み付けとして、*tf-idf* を採用している。

4. 評価実験

システムを評価するため、二つの評価実験を行った。実験項目と目的は以下の通りである。

- 図表画像識別器の精度評価 (実験 1)

目的：テレビ番組の動画から図表画像をどの程度正確に抽出できるかを明らかにする。

- ランキングの妥当性評価 (実験 2)

目的：クエリとして与えられた動向情報と、取得された図表画像のランキングの妥当性を明らかにする。

4.1 図表画像識別器の精度評価 (実験 1)

ここでは、テレビ番組の動画から図表画像をどの程度正確に抽出できるかを明らかにする。

4.2 実験データ

録画したテレビ番組、Web から人手で収集した正例、負例画像、機械学習により構築した円グラフ、棒グラフ、折れ線グラフそれぞれの図表画像識別器を用いる。

4.3 評価方法

円グラフ、棒グラフ、折れ線グラフそれぞれの場合の識別器について、10-fold 交差確認を行い、精度を適合率 (*Precision*)、再現率 (*Recall*) によって評価する。また、テレビ番組の画像フレームに識別器を適用した際の精度も、同様に適合率、再現率によって評価する。その際、識別器にかける図表フレームは、学習に利用していない 2011 年 11 月 1 日から 2011 年 11 月 30 日のテレビ番組からランダムに抽出した。適合率、再現率は、以下の式で算出する。

$$Precision = \frac{tp}{tp + fp} \quad (7)$$

$$Recall = \frac{tp}{tp + fn} \quad (8)$$

ただし、*tp* は正しく抽出できた正解画像数、*fp* は間違っ抽出してしまった不正解画像数、*fn* は抽出できなかった正解画像数である。

4.4 結果・考察

10-fold 交差確認の結果を図 7 に、テレビ番組の図表フレームに適用した際の精度を表 3 に示す。また、HOG+AdaBoost でテレビ番組の図表フレームに適用した結果、正しく抽出できた円グラフの例を図 8 に、間違っ抽出してしまった画像の例を図 9 に、抽出できなかった円グラフの例を図 10 に、正しく抽出できた棒グラフの例を図 11 に、間違っ抽出してしまった画像の例を図 12 に、抽出できなかった棒グラフの例を図 13 に示す。



図 8 正しく抽出できた円グラフの例

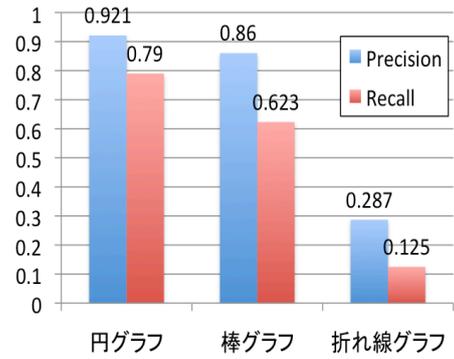


図 7 実験 1 の結果 (10-fold 交差確認)

表 3 実験 1 の結果 (テレビ番組の図表フレームに適用)

項目	円グラフ	棒グラフ
Precision	0.082	0.15
Recall	0.781	0.333
全フレーム数	66348	66348
図表数	32	24
tp	25	9
fp	279	172
fn	7	15



図 9 間違っ抽出してしまった画像の例



図 10 抽出できなかった円グラフの例

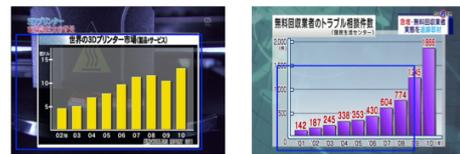


図 11 正しく抽出できた棒グラフの例



図 12 間違っ抽出してしまった画像の例

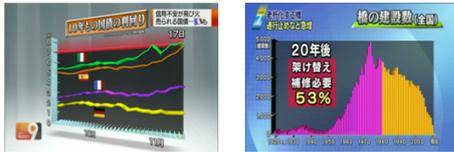


図 13 抽出できなかった棒グラフの例

円グラフについては、ほとんど抽出できているが、円グラフではない丸い形のものを検出してしまっている。これは、円グラフは、データの割合によってグラフの中に引かれる線が変わってしまい、共通した特徴が少ないためと考えられる。JointHOG 特徴量のような局所領域間の共起を利用した特徴量を使用すると、改善できることが予想される。

棒グラフについては、特に適合率が低いことがわかる。これは、学習する正例数が少なく (130 件)、グラフの種類を網羅できていないためと考えられる。

4.5 ランキングの妥当性評価 (実験 2)

ここでは、クエリとして与えられた動向情報と、取得された図表画像のランキングの妥当性を明らかにする。

4.5.1 実験データ

クエリとして、ブログ、2ちゃんねるから人手で取得した 20 件の動向情報を使用する。クエリには、正解が必ず存在するクエリ、正解があるか不明なランダムに取得したクエリの二通りを用意する。

テレビ番組の字幕情報からキーワードを抽出する際、抽出範囲は、図表画像の前後 15 秒、前後 30 秒、前後 60 秒とした。また、キーワードの重み付けは、*tf-idf* と *FLR* 法それぞれで算出し、結果を比較をする。また、*tf-idf*、*FLR* 法のそれぞれで算出した重みの上位 5 件を使用して検索した際の精度、上位 10 件を使用して検索した際の精度、全キーワードを使用して検索した際の精度も比較する。用いたクエリの例を表 4 に示す。

表 4 用いたクエリの例 (実験 2 の実験データ)

クエリ	ジャンル
菅内閣の支持率は NHK の調査で 16 % と先月の調査より 9 % 下がり 一昨年の政権交代以降最低となった一方、不支持率は 11 % 上がって 68 %。	政治
日本の電気料金は OECD 諸国では産業用はイタリアに次いで第二位。	経済
生活保護の受給者数が今年 7 月時点で 205 万人を超え、 過去最多となりました。	医療・福祉
9 割以上は日本に輸出されている、とされている台湾産うなぎ。 今年は 5 月頃から、日本産より高くなっているという。	食
密輸の仕出し国は、昨年まで中心だったアジアに代わり、 押取量の半分をアフリカが占めた。	その他

図表画像に相当する被検索文書は、現在までに取得した 108 件を用いて実験した。108 件のジャンルを表 5 に示す。

4.5.2 評価方法

クエリとして与えられた動向情報の信頼性判断に有効な図表が提示され、信頼性判断に役立つ情報が確認できた場合を正解とする。例えば、「わが国の外食産業全体の売上高は、1997 年をピークに毎年下がり続けています。」というクエリに対し、図 14 が提示された場合、図中から日本の外食産業の売上高の推移が読み取れるため、正解と判断する。また、同クエリに対し、図 15 が提示された場合、図中から日本の外食産業の売上高の推移が読み取れないため、不正解とする。

表 5 被検索文書のジャンル (実験 2 の実験データ)

ジャンル	件数
政治	12 件
経済	36 件
医療・福祉	8 件
環境	8 件
雇用	7 件
健康	8 件
原発	9 件
気象	4 件
食	7 件
その他	9 件



図 14 正解図表の例



図 15 不正解図表の例

評価には上位 10 件での MAP (Mean Average Precision) を使う。MAP は、全正解に対する平均適合率 (*AP*: Average Precision) の平均値によって求められる。平均適合率 (*AP*) とは、検索結果における適合率の平均値であり、上位 k 件での平均適合率は以下の式で表される。

$$AP = \frac{1}{n} \sum_{i=1}^k P_i \quad (9)$$

ここで、 n は正解画像の数であり、 P_i は i 番目の正解画像が検索された時点での適合率である。よって、クエリ N 件での MAP は以下の式で表される。

$$MAP = \frac{1}{N} \sum_{j=1}^N AP_j \quad (10)$$

4.5.3 結果・考察

正解が必ず存在するクエリで検索した場合の結果を図 16、図 17 に示す。

キーワード抽出範囲については、図表画像の前後 15 秒で抽出する場合に最も高い MAP 値が得られることがわかる。この傾向は、用いるキーワード数や、*tf-idf*、*FLR* 法のいずれを用いる場合でも変わらなかった。これは、キーワード抽出範囲を広げすぎると、動向情報に関連しない不要なキーワードが、重み付けした上位にきてしまうためと考えられる。

一方、クエリベクトルに用いるキーワード数については、重み付けした上位 N 件のみを使う場合よりも、全キーワードを用いる場合の方が良好な結果が得られる傾向があることがわかる。これは、重み付けでは上位ではないが、ランキングに有効であるキーワードが存在するためと考えられる。

現時点では、網羅度が低いため、ランダムなクエリで検索した際の実験は行っていない。今後、網羅度がある程度改善された後、実験を行う予定である。

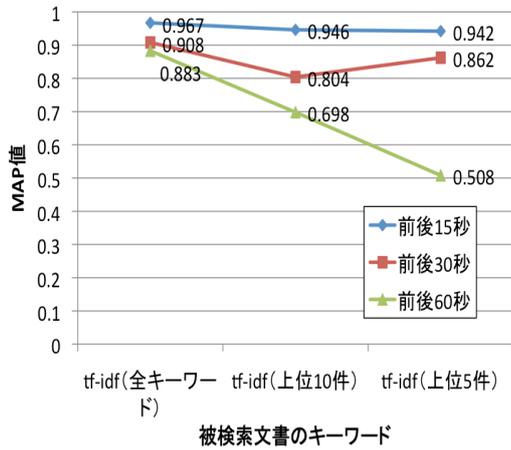


図 16 実験 2 の結果 (tf-idf)

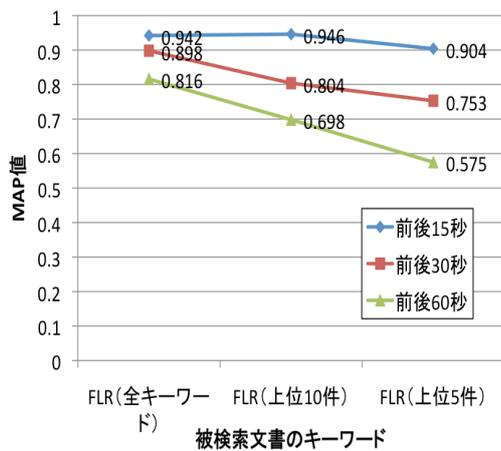


図 17 実験 2 の結果 (FLR 法)

5. まとめと今後の課題

本稿では、Web コンテンツの動向情報を対象に、テレビ番組から得られる関連図表画像を提示することで、該当する動向情報の信頼性判断を効率的に支援するシステムを提案した。予備実験の結果、ランキングの精度で 0.9 以上の MAP 値を確認し、正解が必ず存在するクエリで検索した場合には、クエリとして与えられた動向情報と、取得された図表画像のランキングが可能であることがわかった。これにより、クエリとして入力された動向情報の信頼性判断に有効な図表がテレビ番組中に存在する場合、一定の良好な精度で関連図表をユーザに提示でき、動向情報の信頼性判断を効率よく支援することにつながるという見通しを得た。

今後の課題として、図表画像識別器を改良し、図表画像の抽出精度を向上すること、被験者実験により本システムの有効性を確認すること、および、図表画像の網羅度を上げることが挙げられる。特に、テレビ番組中に出現する図表画像自体が少ないため、網羅度を効率よく向上させることが課題である。ユーザが検索したい動向情報に関連した図表と字幕がある番組は限られているので、どのようにして網羅度を改善するか検討していきたい。

謝 辞

本研究は科研費 20300042 の助成を受けたものである。

文 献

- [1] 宮森恒, 赤峯亨, 加藤義清, 兼岩憲, 角薫, 乾健太郎, 黒橋禎夫: 情報の信頼性分析に向けた評価データおよびプロトタイプシステム WISDOM, 社団法人 情報処理学会 研究報告 NL-180, 2007
- [2] 河原大輔, 乾健太郎, 黒橋禎夫: Web 情報の俯瞰的把握のための主要・対比・対立文の抽出と集約, 言語処理学会第 16 回年次大会, pp.134-137, 2010
- [3] 加藤義清, 乾健太郎, 黒橋禎夫: Web ページの情報発信者の同定とその関係の抽出, 言語処理学会 第 14 回年次大会, pp.737-740, 2008
- [4] 赤峯亨, 加藤義清, 乾健太郎, 黒橋禎夫: Web 情報信頼性判断のためのページ外観情報の抽出, the Institute of Electronics, Information and Communication Engineers Information and Systems Society, Human Communication (2008)
- [5] 中川哲治, 乾健太郎, 黒橋禎夫: 隠れ変数を持つ条件付き確率場による依存構造木の評価極性分類, 第 192 回自然言語処理研究会, 2009
- [6] 山本祐輔, 田中克己: データ対問のサポート関係分析に基づく Web 情報の信憑性評価, 情報処理学会論文誌 Vol.3 No.2, pp.61-79, 2010
- [7] 山内悠嗣, 藤吉弘亘, Bon-Woo Hwang, 金出武雄: アピアランスと時空間特徴の共起に基づく人検出, 画像の認識・理解シンポジウム (MIRU), 2007
- [8] 三井相和, 山内悠嗣, 藤吉弘亘: Joint HOG 特徴を用いた 2 段階 AdaBoost による人検出, 第 14 回 画像センシングシンポジウム SSII08 IN1-06, 2008
- [9] 山内悠嗣, 藤吉弘亘, 山下隆義: Boosting に基づく共起表現による人検出, 画像の認識・理解シンポジウム (MIRU2008), pp.180-187, 2008
- [10] 藤吉弘亘: 局所特徴量の関連性に着目した Joint 特徴による物体検出, 情報処理学会 研究報告 CVIM 166, 2009
- [11] 中川裕志, 湯本紘彰, 森辰則: 出現頻度と接続頻度に基づく専門用語抽出, 2001
- [12] CVIM-Special Lecture
<http://www.vision.cs.chubu.ac.jp/jointhog/>
- [13] Mecab: Yet Another Part-of-Speech and Morphological Analyzer 京都大学情報学研究所-日本電信電話株式会社コミュニケーション科学基礎研究所 共同研究ユニットプロジェクト
<http://mecab.sourceforge.net/>