

【特許請求の範囲】

【請求項 1】

個人がwebサイトへ書き込んだ文字列と、その文字列を書き込んだ時刻情報とが含まれるブログをネットワークを介して収集し、前記ブログの文字列中に出現するタグを取り出し、前記ブログと取り出したタグとを対応づけてタグ出現データベースに格納するタグ取出部と、

前記ブログの文字列中に出現する特徴語をもとに前記ブログで言及されている番組を放送する放送局を推定し、推定した放送局を仮放送局として前記ブログと対応づけてブログデータベースに格納する仮放送局推定部と、

前記タグ出現データベース中に格納されたブログであって所定の時刻範囲内に書き込まれたブログの文字列中に出現するタグの数が所定の閾値を越えた場合、前記ブログデータベースを参照して前記タグを文字列中に含むブログで言及されている番組の仮放送局を集計し、集計結果をもとに前記ブログで言及されている番組を放送する放送局を確定する放送局確定部とを含むことを特徴とする推定装置。

10

【請求項 2】

各放送局が放送する番組の放送時間帯を格納した番組情報データベースを参照し、前記放送局確定部が確定した放送局が放送する番組の中で、前記番組の放送時間帯中に書き込まれた前記ブログの文字列中に出現する前記タグの数をもとに前記タグが示す番組を推定し、前記タグと推定した番組とを対応づけてタグ番組データベースに格納する番組推定部をさらに含むことを特徴とする請求項 1 に記載の推定装置。

20

【請求項 3】

前記タグ出現データベース中に格納されたブログであって所定の時刻範囲内に書き込まれたブログのうち前記番組推定部が番組を推定したタグを含むブログを、前記番組および前記放送局確定部が確定した放送局と対応づけて前記ブログデータベースに格納する番組情報書換部をさらに含むことを特徴とする請求項 2 に記載の推定装置。

【請求項 4】

ネットワークを介して収集したブログの文字列中に、前記番組推定部が番組と対応づけたタグを含む場合、前記ブログが書き込まれた時刻がタグの所定の時刻範囲に収まっていることを条件として、前記タグに対応づけられた番組および前記番組を放送する放送局を前記ブログと対応づけて前記ブログデータベースに格納する番組情報設定部をさらに含むことを特徴とする請求項 2 から 3 のいずれかに記載の推定装置。

30

【請求項 5】

個人がwebサイトへ書き込んだ文字列と、その文字列を書き込んだ時刻情報とが含まれるブログをネットワークを介して収集し、前記ブログの文字列中に出現するタグを取り出し、前記ブログと取り出したタグとを対応づけてタグ出現データベースに格納するステップと、

前記ブログの文字列中に出現する特徴語をもとに前記ブログで言及されている番組を放送する放送局を推定し、推定した放送局を仮放送局として前記ブログと対応づけてブログデータベースに格納するステップと、

前記タグ出現データベース中に格納されたブログであって所定の時刻範囲内に書き込まれたブログの文字列中に出現するタグの数が所定の閾値を越えた場合、前記ブログデータベースを参照して前記タグを文字列中に含むブログで言及されている番組の仮放送局を集計し、集計結果をもとに前記ブログで言及されている番組を放送する放送局を確定するステップとをプロセッサに実行させることを特徴とする推定方法。

40

【請求項 6】

個人がwebサイトへ書き込んだ文字列と、その文字列を書き込んだ時刻情報とが含まれるブログをネットワークを介して収集し、前記ブログの文字列中に出現するタグを取り出し、前記ブログと取り出したタグとを対応づけてタグ出現データベースに格納する機能と、

前記ブログの文字列中に出現する特徴語をもとに前記ブログで言及されている番組を放

50

送する放送局を推定し、推定した放送局を仮放送局として前記ブログと対応づけてブログデータベースに格納する機能と、

前記タグ出現データベース中に格納されたブログであって所定の時刻範囲内に書き込まれたブログの文字列中に出現するタグの数が所定の閾値を越えた場合、前記ブログデータベースを参照して前記タグを文字列中に含むブログで言及されている番組の仮放送局を集計し、集計結果をもとに前記ブログで言及されている番組を放送する放送局を確定する機能とをコンピュータに実現させることを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、推定装置、推定方法、並びにプログラムに係り、特に、収集したブログの特徴語を使用して、ブログが言及している番組を放送する放送局やその番組を推定する技術に関する。

【背景技術】

【0002】

インターネットやブログが一般に普及したことにより、ごく普通のユーザが自身のブログサイトで放送中のテレビ番組などをリアルタイムに実況するということが流行している。特に140文字程度の短いコメントを投稿、閲覧できるようなマイクロブログは、携帯電話からも投稿できる手軽さから、こうした番組の実況という利用が急速に広まっている。なお、本明細書における「ブログ」とは、個人がwebサイトへ投稿したコメントや記事と定義する。

【0003】

このような背景から、ユーザが現在視聴しているテレビ番組の番組情報をユーザへ一覧させることで、その番組に関連したブログ記事やコメントの作成支援を行う発明などもある(特許文献1参照)。

【0004】

また、逆のアプローチとして、ブログ記事の内容からどの番組について実況しているかを推定する手法についても考えられつつある。このようなブログが言及している番組の推定に関して、字幕テキストを使った手法として、「字幕テキストの利用によるブログで引用されたテレビ番組の推定」が挙げられる(非特許文献1参照)。これは、EPG(Electronic Program Guide: 電子番組ガイド)や字幕テキストから特徴語を抽出し、ブログ内の単語と比較することで、テレビ局制約、ジャンル制約を決定し、さらに時制特徴語から時間制約を取り出し、スコアリングを行い、ブログが言及している番組を推定するものである。

【0005】

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2008-124861号公報

【非特許文献】

【0007】

【非特許文献1】及川 孝徳、外4名、「字幕テキストの利用によるブログで引用されたテレビ番組の推定」[online][平成23年9月9日検索]、インターネット<URL:http://db-event.jpn.org/deim2010/proceedings/files/D6-4.pdf>

【0008】

一方、現実のプロガー(ブログを書く人)は、テレビ番組のみを365日実況しているわけではなく、自分の普段の生活の一言や、書籍の感想などを1つのブログサイトに混在して記述する。このようなプロガーの行動様式としては、こういった混在した個々の内容の付箋として、タグ名を付与することが確立している。例えば、日常、書籍、テレビなどのカテゴリに分けて#dairy、#books、#tvなどのタグを付ける。また、

10

20

30

40

50

良く視聴する番組に対しては固有のタグを付与することも良く行われている。例えば、連続ドラマ A #rendora A などの省略語を基調としたものが多い。

【発明の概要】

【発明が解決しようとする課題】

【0009】

しかしながら、前記の提案手法では、字幕テキストが付いていない番組に関しては最初から扱わない（評価でも全て例外にしている）が、現実として番組の半分には字幕テキストが付いておらず、実用的ではなかった。また、ラジオなど元々字幕テキストが存在しない放送に対しては適用が難しかった。

更に、現実の字幕テキストは、シナリオや放送倫理規定に沿った正しい日本語である。一方、ユーザがテレビを見ながらブログに入力する場合は、省略語、スラング、タグが多用されるカジュアルな日本語になりがちである。特にタレント名や番組名などは省略語にされやすいため、思い通りの推定精度が得られない。これを解決する既知の技術として、正しい日本語と、省略語、スラング、タグとの類義語辞書を作成することで、推定精度を高めることも考えられているが、日々未知の単語が現れる類義語辞書をメンテナンスしていくことは高コストである。

また、確かにブロガーの行動様式としてタグを付けることは良く行われているものの、そのタグは決して放送局が指定したのではなく、自然発生的に決まったものであって、出現頻度にもばらつきがあり、番組情報と結びつけることが難しかった。

【0010】

そこで本発明は、EPGや字幕テキストから特徴語を抽出したり、類義語辞書などをメンテナンスしたりすることなく、ブログ等のソーシャルメディアが言及している番組を高精度に推定できるようにすることを目的とする。

【課題を解決するための手段】

【0011】

上記目的を達成するため、本発明のある態様は推定装置である。この装置は、個人がwebサイトへ書き込んだ文字列と、その文字列を書き込んだ時刻情報とが含まれるブログをネットワークを介して収集し、前記ブログの文字列中に出現するタグを取り出し、前記ブログと取り出したタグとを対応づけてタグ出現データベースに格納するタグ取出部と、前記ブログの文字列中に出現する特徴語をもとに前記ブログで言及されている番組を放送する放送局を推定し、推定した放送局を仮放送局として前記ブログと対応づけてブログデータベースに格納する仮放送局推定部と、前記タグ出現データベース中に格納されたブログであって所定の時刻範囲内に書き込まれたブログの文字列中に出現するタグの数が所定の閾値を越えた場合、前記ブログデータベースを参照して前記タグを文字列中に含むブログで言及されている番組の仮放送局を集計し、集計結果をもとに前記ブログで言及されている番組を放送する放送局を確定する放送局確定部とを含む。

本発明の別の態様は、推定方法である。この方法は、個人がwebサイトへ書き込んだ文字列と、その文字列を書き込んだ時刻情報とが含まれるブログをネットワークを介して収集し、前記ブログの文字列中に出現するタグを取り出し、前記ブログと取り出したタグとを対応づけてタグ出現データベースに格納するステップと、前記ブログの文字列中に出現する特徴語をもとに前記ブログで言及されている番組を放送する放送局を推定し、推定した放送局を仮放送局として前記ブログと対応づけてブログデータベースに格納するステップと、前記タグ出現データベース中に格納されたブログであって所定の時刻範囲内に書き込まれたブログの文字列中に出現するタグの数が所定の閾値を越えた場合、前記ブログデータベースを参照して前記タグを文字列中に含むブログで言及されている番組の仮放送局を集計し、集計結果をもとに前記ブログで言及されている番組を放送する放送局を確定するステップとをプロセッサに実行させる。

【発明の効果】

【0012】

本発明によれば、ブログ等のソーシャルメディアが言及している番組を放送する放送局

10

20

30

40

50

を推定することができる。

【図面の簡単な説明】

【0013】

【図1】本発明の実施の形態1における推定装置のブロック図である。

【図2】本発明の実施の形態1におけるフローチャートである。

【図3】本発明の実施の形態1のブログDBデータ構造の例（番組推定前）である。

【図4】本発明の実施の形態1のブログDBデータ構造の例（番組推定後）である。

【図5】本発明の実施の形態1におけるタグ出現DBデータ構造の例である。

【図6】本発明の実施の形態1におけるタグ番組DBデータ構造の例（番組推定前）である。

10

【図7】本発明の実施の形態1におけるタグ番組DBデータ構造の例（番組推定後）である。

【図8】本発明の実施の形態1における特徴語データの例である。

【図9】本発明の実施の形態2における推定装置のブロック図である。

【図10】本発明の実施の形態2におけるフローチャートである。

【発明を実施するための形態】

【0014】

以下、本発明の実施の形態について、図面を参照して詳細に説明する。

（実施の形態1）

【0015】

20

図1は、本発明の実施の形態1に係る推定装置のブロック図である。また、図2は本発明の実施の形態1におけるフローチャートである。

【0016】

推定装置100のブログ収集部101は、インターネット等のネットワーク200を通じてブログ記事を収集し（S1）、図3に示したブログDB（Data Base；データベース）102に投稿固有ID、本文、投稿時刻を登録する（S2）。なお、この時点では、後述する仮放送局や確定放送局、番組情報（タイトル、放送局名、番組詳細など）は不明であるため、空のままである。なお、本明細書における「ブログ」とは、個人がwebサイトへ1回に投稿した（書き込んだ）コメントや記事を1つの単位とし、その投稿されたコメントや記事を構成する文章情報（文字列）と、そのコメントや記事を投稿した時刻情報とが含まれているものとする。

30

ここで、投稿固有IDは個々のブログ記事に特有のID（Identification）である。本発明では特に投稿固有IDの書式は限定しない。また、ブログ記事を収集するブログサイトについても、特に限定するものではない。

【0017】

次に、タグ取出部103はブログの本文から任意のタグを取り出す（S3）。本実施例では#rendoraAのように“#”から始まる英数字をタグとして取り出しているが、このタグの書式については、本発明で特に限定するものではなく、予め定められたルールに従って機械的に自動取出しできる、ブログ固有の書式（HTMLタグのうち特定の書式で囲まれたテキストなど）やユーザの行動様式に従った書式であれば、どのような書式でも構わない。

40

【0018】

続けて、タグ取出部103は取り出したタグをブログの投稿固有IDおよび投稿時刻と対応づけてタグ出現DB104に登録する（S4）。タグ出現DB104は図5に示したデータ構造を持つ。なお、図5では、後述する仮の放送局名を推定する際に使用する特徴語データは除いた例が示されている。

更に、タグ取出部103は、取り出したタグがタグとそのタグから推定された番組情報の関連付けを示したタグ番組DB105にまだ登録されていない新しいタグであるかを判断し（S5）、新しいタグであるならばタグ番組DB105に登録する（S6）。タグ番組DB105は図6に示したデータ構造を持つ。なお、この時点では確定時刻（タグと番組

50

組情報の関連付けが決定した時刻)と番組情報、および後述する確定放送局は不明であるため、空である。

【0019】

一方、仮放送局推定部106は、ブログ収集部101が収集したブログの本文から仮の放送局を推定し(S7)、推定された仮放送局をブログDB102に一時保存する(S8)。

本発明の実施の形態では、この仮放送局の推定に、図8の例に示したような放送局ごとに予め用意した特徴語データを用いる。この特徴語データは、ブログ本文中に出てくる用語であり、必ずしも上述したタグとは限らない。特徴語データには優先度スコアを予め決めておく。例えば、図3の投稿固有ID=06565406541では、ブログ本文の「Bテレビ」という文字列のみにマッチするので「放送局B」を仮放送局と推定する。この方法は、従来技術のように、EPGや字幕テキストを形態素解析して特徴語を取り出すような必要がなく、推定に掛かる計算コストを大幅に抑えることができる。

仮放送局を推定する精度を上げるために、図8のように特徴語データ毎に優先度のスコア付けを行い、ブログDB102の仮放送局に一時保持し、一つのブログ本文が複数の放送局の特徴語データにマッチする場合(例えば「放送局Aの衛星放送を見てみなよ」というブログ本文は、放送局Aと放送局Aの衛星放送との両者にマッチする)、放送局ごとに合計し、最も仮合計スコアが高い放送局を仮放送局と推定してもよい。この優先度スコアも、特徴語同様、予め定められた値であり、各放送局向けに一度設定した静的なデータである。マッチした仮合計スコアは、ブログDB102の所定のカラムへ一時保存する。

【0020】

この段階での仮放送局はあくまでも仮の放送局であり、マッチングする仮放送局が複数あっても、また、一つもマッチせずに仮放送局が不明であっても構わない。例えば、図3の投稿固有ID=06565406567では、ブログ本文から放送局が推定できないため、対応する仮放送局の欄は「?(不明)」のままである。

【0021】

次に、図5に示したように、放送局確定部107は、定期的にタグ出現DB104を予め定めた規定の時刻範囲Ra(例えば10分前~現在時刻など)で区切り、その時刻範囲Raに任意のタグTが予め定められた閾値回(例えば50回など)以上出現するかを判断し(S9)、タグTが回以上出現する場合に、後述する方法で放送局を確定する。ここで規定の時刻範囲Raとは、タグと放送局との対応関係を求める際に基準とする放送局推定基準時刻範囲である。また、閾値は、タグと放送局とを対応づけるか否かを決定するために参照する、放送局決定基準値である。

なお、本実施例では説明の簡単化のために、前記規定の時刻範囲を10分間に固定した例を説明しているが、定期的に時刻を監視し、現在放送中の番組情報から、番組の開始時刻と終了時刻を取り出し、現在時刻が終了時刻に重なった時点で、時刻範囲Raを決定することで、時刻範囲を可変にしてもよい。

【0022】

ここでは、現在時刻22:05において、図5で示した規定の時刻範囲Ra(21:55:00 22:05:00)の区間に#prog1というタグTが回以上出現したとする。

放送局確定部107は、タグ出現DB104に記録した投稿固有IDを元にブログDB102から#prog1というタグが含まれている最も古い投稿時刻(20:50:22)から最新の投稿時刻(22:02:20)まで時刻範囲Rb内のブログリストLbを取得する(S10)。放送局確定部107は次に、このブログリストLbから仮放送局の出現回数を集計してランキングを作成し、最も出現回数の多い仮放送局を#prog1が示す確定放送局と特定する(S11)。

なお、本実施例においてブログDB102は、ブログ収集部101が収集した過去1週間分のブログを格納し、それよりも過去のブログは削除しているものとする。また、説明の簡単化のために、タグTが含まれている最も古い投稿時刻をそのまま時刻範囲Rbに採

10

20

30

40

50

用したが、最も古い時刻は、最新の投稿時刻と同一日時であるといった制限を加えて、同一番組が異なる地方局で日時をずらして放送するといった編成に対処してもよい。

【 0 0 2 3 】

より多くの例を示す。例えば、時刻範囲 R b (2 0 : 5 0 : 2 2 2 2 : 0 5 : 0 0) の区間に、全体で 3 0 5 件のタグ T を含む投稿があり、仮放送局ごとの集計結果が

1 位 : 放送局 A = 2 9 6 件 (合計スコア 4 8 5 . 0)

2 位 : 放送局 E = 6 件 (合計スコア 8 . 5)

3 位 : 放送局 F = 1 件 (合計スコア 2 . 0)

不明 = 2 件 (不明率 0 . 7 %)

であった場合には、「放送局 A」を確定放送局とする。この方法により、一つ一つのブログ記事から推定した仮放送局のばらつきや誤差を抑えることができ、推定精度を格段に高めることができる。

また、本実施例では説明の簡単化のために必ず放送局が確定するようになっているが、より精度を高めるために、仮放送局の分布を統計的に判定し、棄却条件に一致（ここで棄却条件は、例えば、1位と2位の合計優先度スコアに殆ど差がない場合、不明率が全体の投稿件数に対して著しく大きい（例えば不明率30%以上など）場合を指す）した場合には、タグ T が特定の放送局や番組情報を示すものではないとして、番組情報の推定には用いないと判断してもよい。

【 0 0 2 4 】

前記の方法で確定放送局が特定できたならば、次に番組推定部 1 0 8 は番組情報 D B 1 0 9 から前記時刻範囲 R b に該当する確定放送局の全ての番組候補を取得する (S 1 2) 。ここで、番組情報 D B 1 0 9 は少なくとも放送局名、放送時間、タイトル、番組詳細などの情報を蓄積したデータベースであるが、これらの情報の取得部について本発明では特に限定しない。ネットワーク経由で取得した情報でも、放送波に含まれる電子番組データから取得したもので、その他の方法で取得したもので構わない。

【 0 0 2 5 】

先ほどの例では、前記時刻範囲 R b (2 0 : 5 0 : 2 2 - 2 2 : 0 5 : 0 0) に該当する放送局 A の番組候補として、2 0 : 4 5 - 2 1 : 0 0 に放送の「番組 5」と、2 1 : 0 0 - 2 2 : 0 0 に放送の「番組 6」、2 2 : 0 0 - 2 2 : 4 5 に放送の「番組 1」の3つの番組が得られる。

更に、前記プログラリスト L b の投稿時刻を元にして、それぞれの番組放送時間中の投稿件数を集計すると、ユーザがリアルタイムに番組を実況しているという行動様式から、以下のように特定の番組についての投稿件数が非常に大きくなる。

「放送局名：放送局 A

タイトル：番組 5

放送時間：2 0 : 4 5 - 2 1 : 0 0

前記プログラリスト L b 中の投稿件数：1 件」、

「放送局名：放送局 A

タイトル：番組 6

放送時間：2 1 : 0 0 - 2 2 : 0 0

前記プログラリスト L b 中の投稿件数：5 件」、

「放送局名：放送局 A

タイトル：番組 1

放送時間：2 2 : 0 0 - 2 2 : 4 5

前記プログラリスト L b 中の投稿件数：2 9 9 件」。

この結果から、番組推定部 1 0 8 は、前記プログラリスト L b 中の投稿件数が最も多い「番組 1」がタグ # p r o g 1 が示す該当番組であるとして推定する (S 1 3) 。ここでも、複数の番組に関連するブログのばらつきや誤差を抑えることで、推定精度を高めている。

【 0 0 2 6 】

10

20

30

40

50

タグ T が示す番組が推定できた時点で、番組推定部 108 は、タグ番組 DB 105 に番組情報として番組情報 DB 109 から取得した「番組 1」に関する情報を登録し、確定時刻として現在時刻の 22:05 を登録することで、タグ T と番組情報との関連付けを完了する (S14)。

【0027】

最後に、番組情報書換部 110 が前記プログラリスト Lb のうち、番組情報が空であるものに関して、推定した番組情報を元にプログ DB 102 の番組情報および確定放送局を書き換える (S15)。これにより、S7 の段階では不明であった番組情報や誤って推定された確定放送局も正しい情報となる。

例えば、図 3 に示した投稿固有 ID = 06565406567 は、ログ本文からは仮放送局を推定できていなかったが、前記手法によって、図 4 に示したように放送局が確定されている。また、図 4 および図 7 は同様にして番組情報が確定した状態を示している。

図 4 の投稿固有 ID = 06565406542 は #dairy というタグを含んでいるものの、そのタグが特定の番組と関連づけられる結果が得られなかったため、番組を引用しているログではないと判別できる。

【0028】

以上のように、実施の形態 1 によれば、EPG や字幕テキストから特徴語を抽出したり、類義語辞書などをメンテナンスしたりすることなく、ログ等のソーシャルメディアが言及している番組を高精度に推定できるという効果を得ることができる。

(実施の形態 2)

【0029】

図 9 は、本発明の実施の形態 2 に係る推定装置のブロック図である。また、図 10 は本発明の実施の形態 2 におけるフローチャートである。実施の形態 1 の推定装置 100 に、番組情報設定部 111 を加えたブロック構成になっている。

【0030】

図 10 のフローチャートにおいて、ステップ S1 からステップ S4 まで、および、ステップ S6 からステップ S15 までは実施例 1 と同様であるため、説明は省略する。

【0031】

タグ取出部 103 は、タグとそのタグから推定された番組情報の関連付けを示したタグ番組 DB 105 にまだ登録されていない新しいタグであるかを判断し (S5)、新しいタグであるならばタグ番組 DB 105 に登録する (S6)。

既にタグ番組 DB 105 に登録されているタグである場合、番組情報設定部 111 は、そのタグに関連付けられている番組情報があるかどうかを判断し (S16)、番組情報がある場合には、ログの投稿時刻がタグ番組 DB 105 にある確定時刻 + 閾値 以内であるかどうかを判断し (S17)、範囲内であればタグが同一の番組情報を示していると推定し、プログ DB 102 の番組情報および放送局リストを書き換える (S18)。ここで閾値 は、タグと番組情報との対応関係を求める際に基準とする番組推定基準時刻範囲である。

【0032】

例えば、予め定められた閾値 = 2 時間である場合に、図 3 に示した投稿固有 ID = 06565406721 では、ログ本文と特徴語データとのマッチングからは仮放送局が得られないが、タグ #prog1 を含んでおり、かつ、投稿時刻 22:43:20 がタグ #prog1 の確定時刻 22:05:00 + 閾値 (2 時間) 以内であるため、番組情報設定部 111 により、「番組 1」であると直ちに推定される。その結果、図 4 に示したように投稿固有 ID = 06565406721 の番組情報および放送局が設定される。

【0033】

確定時刻 + 閾値 による制限は、同一の番組が異なる放送局で異なる時間に放送された場合に、古い関連付けの番組情報を強制しないためのものである。例えば、放送局 F と放送局 G で同一の番組が異なる時間帯で放送されることがある。しかしながら、放送局名が異なるため、番組情報は完全には一致しない。このとき、後に放送された番組に対して、

10

20

30

40

50

新しい番組情報を関連付けるためにこのような制限が必要となる。このように、閾値の値は各局で放送される番組情報を勘案して実験により定めればよい。

【0034】

また、同様の考えから、タグ番組DB105を定期的に確認し、確定時刻が現在時刻 - 閾値以前のタグを削除することで、タグと番組情報の関連づけを新しいものに更新していくことも可能である。

【0035】

以上のように、実施の形態2によれば、番組に関連するタグTは設定されているが、既知の特徴語テーブルからは仮の放送局が推定できないプログテキスト（即ち、従来は活用されずに捨てられていたプログテキスト）の番組を推定するという効果を得ることができる。

10

【0036】

以上、本発明を実施の形態をもとに説明した。実施の形態は例示であり、それらの各構成要素や各処理プロセスの組み合わせにいろいろな変形例が可能なこと、またそうした変形例も本発明の範囲にあることは当業者に理解されるところである。

【0037】

（変形例1）

上記の説明では、放送局確定部107が、タグ出現DB104に記録した投稿固有IDを元にプログDB102から特定のタグに関するプログリストLbを取得し、取得したプログリストLbから仮放送局の出現回数を集計してランキングを作成し、最も出現回数の多い仮放送局をそのタグが示す確定放送局と特定する場合に説明した。放送局確定部107によるタグと放送局との対応の付け方は、出現回数の最大値に基づく方法に限らない。放送局確定部107は、確定放送局を求める際に、単にプログリストLbに出現する仮放送局の出現回数を集計するのみならず、その集計結果をさらに解析することで確定放送局を求めてもよい。以下、放送局確定部107によるタグと放送局との対応付けの、別の例について説明する。

20

【0038】

放送局確定部107は、まずプログリストLbから推定される各仮放送局の数の時系列的な変化を算出する。時系列的な変化は、例えば時刻を横軸とし、各時刻において投稿されたプログから推定される仮放送局の数を縦軸とするグラフによって表現される。続いて放送局確定部107は、算出した時系列的な変化の時間微分を算出し、微分値の極大値が最も大きくなるグラフを持つ仮放送局を、そのタグに対応する放送局とする。このように時系列的な変化の微分値を求めることは、プログ中の瞬間的な盛り上がりを評価することに相当する。これにより、ある番組のスタート時点や番組中の人気コーナーの放映中など、番組の進行状況に応じた盛り上がりの変化をその放送局の決定に反映させることができる。

30

【0039】

（変形例2）

上記の説明では、番組推定部108は、プログリストLb中の投稿件数が最も多い「番組」を、タグが示す該当番組であるとして推定する場合について説明した。番組推定部108によるタグと番組との対応の付け方は、投稿件数の最大値に基づく方法に限らない。以下、番組推定部108によるタグと番組との対応付けの、別の例について説明する。

40

【0040】

番組推定部108は、プログリストLb中の投稿された番組の数を集計し、その数を番組の放送時間で規格化した規格化番組数をもとにタグと番組とを対応づけてもよい。より具体的には、番組推定部108は、集計した番組の数をその番組の放送時間で割ることにより、単位放送時間あたりの番組数を求める。一般に、放送時間の長い番組の方が、放送時間が短い番組と比較して、プログに投稿される数が増えると考えられる。例えば、放送時間が10分の番組の後に放送時間が3時間の番組が放送されることを考えると、プログに投稿される数の総和は、放送時間が3時間の番組の方が多くなると考えられる。したが

50

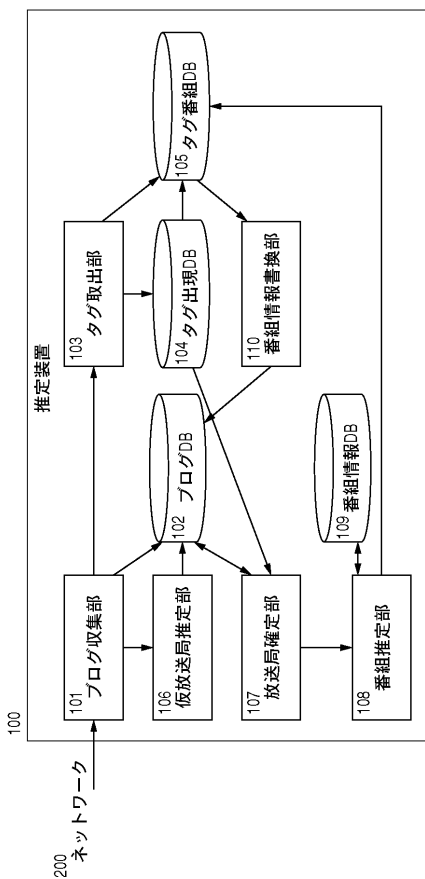
って、番組推定部 108 が単位放送時間あたりの番組数をもとにタグと番組とを対応づけることにより、放送時間による投稿数の相違を低減し、番組推定精度を向上することができる。

【符号の説明】

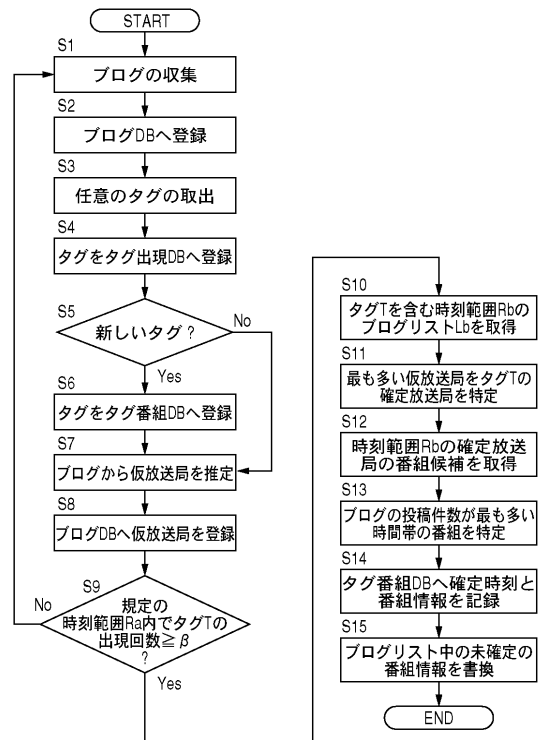
【0041】

100 推定装置、 101 ブログ収集部、 102 ブログDB、 103 タグ取出部、 104 タグ出現DB、 105 タグ番組DB、 106 仮放送局推定部、 107 放送局確定部、 108 番組推定部、 109 番組情報DB、 110 番組情報書換部、 111 番組情報設定部、 200 ネットワーク。

【図1】



【図2】



【 図 3 】

投稿固有ID	本文	投稿時刻	仮放送局 (仮合計スコア)	確定放送局	番組情報 タイトル、放送局名、番組詳細
06565406125	今日は番組があるね。#STA #prog 1	2010/11/11-20:50:22	放送局A [2.0]	? (不明)	? (不明)
06565406540	アングルがもうすぐ始まるよ。#sa #prog 1	2010/11/11-21:58:13	放送局A [1.0]	? (不明)	? (不明)
06565406541	今、Bテレで見えます #STB #prog 2	2010/11/11-22:00:59	放送局B [1.0]	? (不明)	? (不明)
06565406542	帰宅中、なう #daily	2010/11/11-22:02:06	? (不明)	? (不明)	? (不明)
06565406567	今回は銀ブラ? #prog 1	2010/11/11-22:02:20	? (不明)	? (不明)	? (不明)
06565406568	とても暇です。 #daily	2010/11/11-22:05:52	? (不明)	? (不明)	? (不明)
06565406721	今回のゲストは? #prog 1	2010/11/11-22:43:20	? (不明)	? (不明)	? (不明)
06565406722	今、Cテレの新番組見えます #prog 3	2010/11/11-22:50:12	放送局C [0.5]	? (不明)	? (不明)
06565406723	え!?ほんと? #STD #prog 4	2010/11/11-22:51:33	放送局D [2.0]	? (不明)	? (不明)
06565406724	番組終わってた・・・ #STA #prog 1	2010/11/11-22:53:15	放送局A [2.0]	? (不明)	? (不明)

【 図 4 】

投稿固有ID	本文	投稿時刻	仮放送局 (仮合計スコア)	確定放送局	番組情報 タイトル、放送局名、番組詳細
06565406125	今日は番組があるね。#STA #prog 1	2010/11/11-20:50:22	-	放送局 A	番組 1、放送局 A、...
06565406540	アングルがもうすぐ始まるよ。#sa #prog 1	2010/11/11-21:58:13	-	放送局 A	番組 1、放送局 A、...
06565406541	今、Bテレで見えます #STB #prog 2	2010/11/11-22:00:59	-	放送局 B	番組 2、放送局 B、...
06565406542	帰宅中、なう #daily	2010/11/11-22:02:06	? (不明)	? (不明)	? (不明)
06565406567	今回は銀ブラ? #prog 1	2010/11/11-22:02:20	-	放送局 A	番組 1、放送局 A、...
06565406568	とても暇です。 #daily	2010/11/11-22:05:52	? (不明)	? (不明)	? (不明)
06565406721	今回のゲストは? #prog 1	2010/11/11-22:43:20	-	放送局 A	番組 1、放送局 A、...
06565406722	今、Cテレの新番組見えます #prog 3	2010/11/11-22:50:12	-	放送局 C	番組 3、放送局 C、...
06565406723	え!?ほんと? #STD #prog 4	2010/11/11-22:51:33	-	放送局 D	番組 4、放送局 D、...
06565406724	番組終わってた・・・ #STA #prog 1	2010/11/11-22:53:15	-	放送局 A	番組 1、放送局 A、...

【 図 5 】

投稿固有ID	タグ	投稿時刻
06565406125	#prog 1	2010/11/11 - 20:50:22
06565406540	#prog 1	2010/11/11 - 21:58:13
06565406541	#prog 2	2010/11/11 - 22:00:59
06565406542	#daily	2010/11/11 - 22:02:06
06565406567	#prog 1	2010/11/11 - 22:02:20
06565406568	#daily	2010/11/11 - 22:05:52
06565406721	#prog 1	2010/11/11 - 22:43:20
06565406722	#prog 3	2010/11/11 - 22:50:12
06565406723	#prog 4	2010/11/11 - 22:51:33
06565406724	#prog 1	2010/11/11 - 21:53:15

時刻範囲Ra 21:05 - 22:05現在

時刻範囲Rb 20:50 - 22:05現在

【 図 6 】

タグ	確定時刻	確定放送局	番組情報 タイトル、放送局名、番組詳細
#prog 1	? (不明)	? (不明)	? (不明)
#prog 2	? (不明)	? (不明)	? (不明)
#daily	? (不明)	? (不明)	? (不明)
...
#prog 3	? (不明)	? (不明)	? (不明)
#prog 4	? (不明)	? (不明)	? (不明)

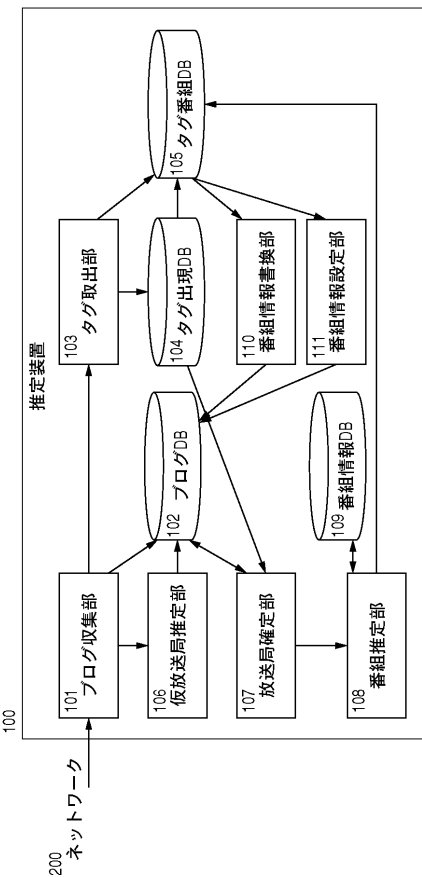
【 図 7 】

タグ	確定時刻	確定放送局	番組情報 番組名、放送局名、番組詳細
#prog 1	2010/11/11-22:05:00	放送局 A	タイトル、放送局名、番組詳細 番組 1、放送局 A、...
#prog 2	2010/11/11-22:15:00	放送局 B	番組 2、放送局 B、...
#daily	? (不明)	? (不明)	? (不明)
⋮	⋮	⋮	⋮
#prog 3	2010/11/11-22:25:00	放送局 C	番組 3、放送局 C、...
#prog 4	2010/11/11-22:25:00	放送局 D	番組 4、放送局 D、...

【 図 8 】

仮放送局	特徴語	スコア
放送局 A	放送局 A	2.0
	A 放送局	2.0
	A テレビ	1.0
	テレ A	1.0
	A	0.5
放送局 A 衛星	放送局 A 衛星	2.0
	衛星 A	1.0
	BSA	0.5
	⋮	⋮
⋮	⋮	⋮
放送局 B	放送局 B	2.0
	B 放送	1.0
	B テレビ	1.0
	B テレ	0.5
	⋮	⋮
⋮	⋮	⋮

【 図 9 】



【 図 10 】

