

討論型ニュースグループを対象とする知的ニュースリーダーの開発

井佐原均、小作浩美、内元清貴
(通信総合研究所)

ネットワーク上の討論型ニュースグループを対象とする知的ニュースリーダー HISHO を開発している。本システムはユーザが興味を持った記事を探索のキーとし、その記事にいたる議論の流れや関連する内容の記事をユーザに提示することにより、効率のよい情報収集を可能とするものである。本稿では、討論型ニュースグループの特徴、本知的ニュースリーダーの基本的な考え方、具体的な機能とシステム構成について述べる。本システムは、プロトタイプが完成し、現在実用システムに向けて開発を進めている。

Intelligent Network News Reader for Discussion Type News Groups

Hitoshi Isahara, Hiromi Ozaku and Kiyotaka Uchimoto
Communications Research Laboratory

We are developing an Intelligent Network News Reader which extracts news articles for the user. In contrast to ordinary information retrieval and abstract generation, this method utilizes an "information context" to select articles from newsgroups on the internet.

The salient feature of this system is that it retrieves articles dynamically, adapting to the user's interests, not classifying them beforehand. Since this system measures the semantic distance between articles, it is possible to refer to the necessary information without being constrained within a particular news group.

We finished our prototype of the Intelligent Network News Reader in March 1997 and are developing its practical version.

1. はじめに

我々は、個々のユーザがネットニュースの情報を効率的かつ有効に利用することをねらいとして、ユーザが興味をもった記事に対して関連記事を検索する上での問題点を解決し、話題の関連性に基づいて記事を読みすすめるような知的ニュースリーダーを開発している [1, 2]。ここでは、従来の情報検索や要約生成とは異なり、記事を一種の対話テキストとみなして、記

事の話の流れを「情報文脈」として扱う。これにより、ユーザは、興味を持った記事から、その記事の話題に関する情報文脈を遡り、必要な情報を過不足なく得ることができる。

知的ニュースリーダーは平成8年度にプロトタイプが開発され、平成9年度に実用システム第1版が、平成11年度末までに実用システムの公開版が完成する予定である。

2. ネットワークニュースの特徴

ネットワークニュースには大きく分けて、「新聞記事型 (newswire-like newsgroups)」と「討論型 (newsgroups for discussion among users)」の二つの種類がある。新聞記事型は fj.meeting に代表されるように何かの事実を通知するタイプのものである。もちろん、このようなニュースグループ中の記事も大変に有効ではあるが、その記事数は多くはない。さらに最近では、この種のアナウンスは WWW のホームページでなされることが多い。一方、討論型には、その他大部分のニュースグループが含まれる。質問に対して回答が行われ、さらにコメントがなされ、といった過程を経るものである。新聞記事型の記事はおおむね自己完結型であり、キーワードを利用する単純な手法でも同じ内容を取り扱う先行記事を検索することが可能である。しかしながら、討論型の記事は内容的にも参照的にも自己完結ではない。このため、検索において何らかの技巧が必要となる。現実のネットワークニュースにおける記事の量、情報の量を考えるに、討論型のニュースグループを無視することはできない。むしろ、討論型のニュースグループの方に有益な情報が多いと言えよう。

これまでの情報検索は、キーワードを用いた全文検索や、そこからの要約生成・情報抽出が研究の中心であった。しかし、ユーザの実際の利用を考えた場合、このような手法は討論型のネットワークニュースに適した方法とは言えない。自分が興味を持った内容については、要約してしまっただけでは重要な情報が落ちる可能性があり、むしろ適切なスレッド (議論の流れ) を選んで、その全文を提示することが望ましい。また、毎日は見られないが、時間がある時には、ニュース (群) を眺めて見る、といったニーズに対応できるシステムが必要である。

3. 知的ニュースリーダー HISHO

我々は、人間の知的創造性の向上を支援する環境開発の一環として、情報の密度を上げるた

めのスクリーニング技術に注目し、知的ニュースリーダーの開発を行っている。本研究ではまずネットワークニュース上の情報の効率的利用を妨げている問題点を明確にし、問題点の解決を試みる。さらに記事の話題の一貫性に注目することにより必要な情報を取り出す方法を提案する。我々のシステムの特徴は、ユーザが興味を持った記事を情報検索のキーとして用いること、キーワード抽出等の自然言語処理技術を適用して記事の間の意味的距離を計算すること、ネットワークニュースをある種の対話テキストと考え、そのヒューリスティックスを用いること、である。

本システムの想定する典型的な利用法は次のようなものである。

ユーザは多忙で、毎日ニュース記事を読むことができない。数日ぶりに少し時間ができて、膨大な未読記事の中から、最近のニュース記事をいくつか読み、その中に非常に興味のある記事を見つけた。ユーザは、これまでの、この話題についての議論の流れ、あるいは関連する投稿を知りたいと思った。さて、ニュースリーダーはどうやってユーザを支援するのか。。。。。

忙しい人のためのニュースリーダーは単に最近のニュースの要約を提示するだけでなくニュース記事の内容に沿って適切な流れを選択するものでなければならない。そこで、我々は記事中の単語の統計的分布等によって決定される「情報文脈」の概念を提案する。この「文脈」を用いることにより、ユーザは、その記事に適切なタイトルや参照関係がない場合にも、話題の流れを追うことができる。

記事検索や要約生成に関する決定を行うためにはキーワードやニュースのサブジェクトフィールド中のタイトルをシステムに与えるだけでは十分ではない。従来のニュースリーダーは、これらの情報を用いて記事を分類しているが、たとえばタイトルはしばしば記事の内容とは異なるので、そのような単純なスクリーニング手法では不必要な記事を数多く抽出してしまう。

一般にキーワード法はどのような情報がほしいかが明確に分かっている場合、あるいは、シソーラスのようなキーワード階層について十分な知識がある場合には有効である。人が基礎的な概念から新しい概念を創造しようとしている過程においては、適切なキーワードを選ぶことは不可能である。人間の概念形成は基礎段階から、関連するニュースを取り出す過程における思考過程を経て、その目的と結果が明確化されていくものと思われる。我々の知的ニュースリーダーは、ネットワークニュースの検索効率を向上するとともに、人間の知的活動を支援するツールの一端を担うことを期待している。

ここでは、検索のキーはユーザが直感によって決定するキーワードやタイトルではなく、関連する記事そのものである。言い換えると、このシステムは曖昧キーを用いた情報検索を可能とする。ユーザは正確なキーワードや記事のタイトルを入力する必要はなく、単に興味を持った記事を指定すればよい。システムが、その記事に関連する記事を見つけてだしてくれる。

システムの実際の利用手順は概略、以下のようになる。

- 1) ユーザが記事群中に興味のある記事を見つける。
- 2) サブジェクト、リファレンス情報を用いてリファレンスツリーを成長させる。
- 3) リファレンスツリー中の各枝分かれについて、関連の深い枝分かれか、そうでないかを判定
- 4) 関連の深いものについてはさらにリファレンスツリーを成長させていく。
- 5) 元記事の意味的特徴をキーとして、リファレンスで繋がっていない関連記事をも抽出し、ツリー構造を成長させる。

このシステムの主要な特徴は、記事をあらかじめ分類することなく、ユーザの興味に沿って、動的に検索することにある。本システムは辞書を用いずに記事間の意味的距離を測定するため、特定のニュースグループに範囲を限らずに必要な情報を抽出することが可能である。

4. 知的ニュースリーダーの機能

知的ニュースリーダーのシステムはその機能によって、ニュースリーダー部と言語処理部の二つに分割される。システム全体としてクライアント・サーバタイプの形態をとる本システムにおいて、クライアントに位置づけられるニュースリーダー部は知的ニュースリーダーが実装した情報検索機能のユーザに対するインタフェースを提供する。言語処理部はサーバとして動作し、ネットワークニュースへの実際の検索処理を行う。

知的ニュースリーダーシステムは、ネットワークニュースを複数の話者による対話テキストとして捉え、以下のような手法を用いて記事の検索を行う。

ニュース記事による対話の流れの構成・復元

ニュース記事はメッセージIDを持ち、各々の記事が参照する記事のメッセージIDをReferencesヘッダという形で保持することにより、話題の参照関係を保存している。このReferencesヘッダを辿ることにより、ニュース記事は一つの記事を先頭にした、木構造を持ったグループ(リファレンスツリー)に分類することが可能である。しかし現実には(1)ReferencesヘッダのメッセージIDを持つ記事が、記事転送上の遅延によりニュースサーバに届いていない。(2)恣意的な編集や、その他の理由により、Referencesヘッダが破損している。といった理由により意味的に完全なリファレンスツリーをニュースリーダーにおいて構成することには困難がある。これに対し、知的ニュースリーダーは、ダミー記事の挿入や、記事の日付情報などを用いた参照の前後関係の推定などの手法を用いて、リファレンスツリーをできる限り復元する機能を持つ。

ユーザが興味を持った話題の抽出

ニュース記事においても通常の対話と同様に、話題の拡散や逸脱が起り、しばしば、同一のリファレンスツリー内に複数の話題が併存する。知的ニュースリーダーはリファレンスツリー内での話題の転換点を検出し、ユーザがある話題に興味を持ったときに読むべき記事の範囲を提示

する機能を持つ。話題転換点検出手法は文献 [3] において提案された二つの手法を組み合わせる。

抽出された話題をキーとした、類似した話題の検索

ネットワークニュースではしばしば同じ、または類似の話題を扱った記事が投稿される。知的ニュースリーダはユーザがある記事を読んでそこに含まれる話題に興味を引かれたとき、同じ、または類似の話題を扱ったリファレンスツリー

を検索する機能を持つ。この検索は複数の話題間の類似度をもって行うが、類似度は、話題転換点で分割されたリファレンスツリーの部分集合である記事群から抽出したキーワードとその頻度・スコアからベクトルを求め、それをもとに、他のリファレンスツリーから同様に求めたベクトルとの内積を計算することによって求める。キーワードの抽出とそのスコア付けには文献 [4] の手法を用いる。今後は神経回路網モデルを用いた記事間の類似度計算を行う予定である。

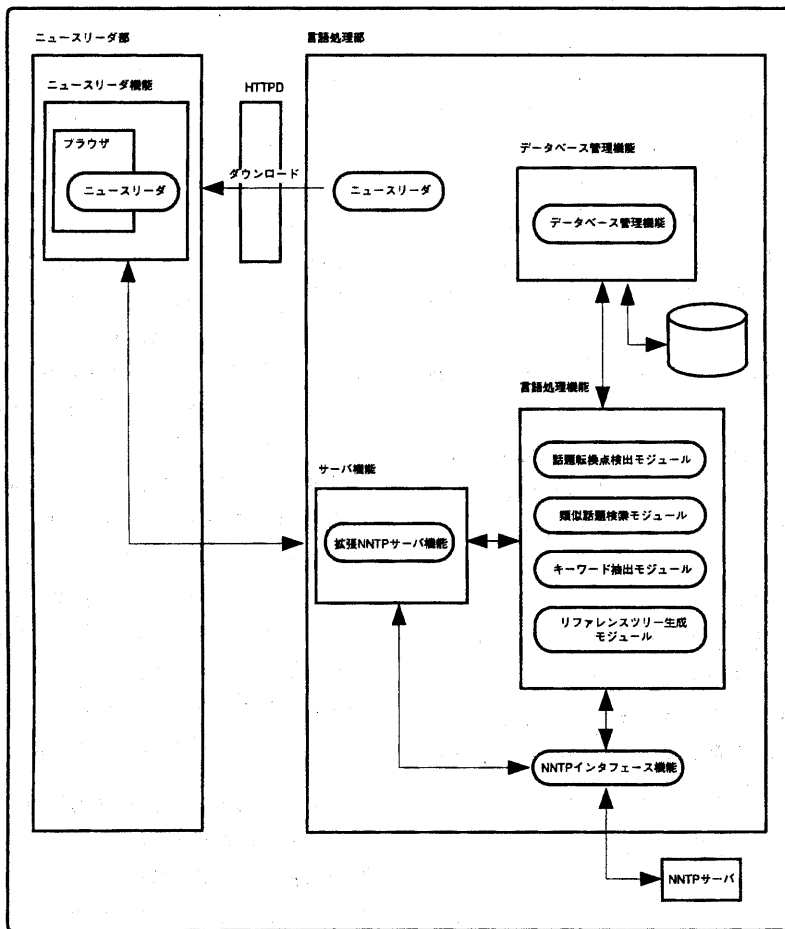


図1 知的ニュースリーダのシステム構成

5. 知的ニュースリーダーの構成

知的ニュースリーダー・プロトタイプシステムは図1に示すように、(1) ユーザのデスクトップの計算機で動作し、グラフィカルユーザインタフェースを提供するニュースリーダー部と、(2) ニュースリーダー部のサーバとして動作するサーバ機能、サーバ機能が受け付けたリクエストを実際に処理する言語処理機能、キャッシュデータベースを管理するデータベース管理機能を含む言語処理部からなる。

言語処理部はニュースリーダー部のサーバとして機能し、ニュースリーダー部のリクエストに対してサービスを提供する。ネットニュースに対する分析や統計的な処理はすべてこの部分で行われる。言語処理部は、その核となる機能である言語処理機能のほかに、ニュースリーダー部からのリクエストに対する窓口となるサーバ機能、処理の高速化のため、言語処理機能が必要とするデータをキャッシングし、それをデータベースとして管理するデータベース管理機能を持つ。

言語処理機能はデータベース管理機能、またはNNTPインタフェースを介してネットニュースの情報を取得し検索処理を行う。この機能は図に示す4つのモジュールによって実現される。

データベース管理機能は、言語処理機能の各モジュールがNNTPサーバより取得したデータや、検索処理の結果、また、検索過程で出力されたキーワードなどのデータをデータベースとして管理する。このデータベースはキャッシュとして機能し、言語処理機能のモジュールがコストの高い計算を繰り返さない。

サーバ機能はニュースリーダー部に対する言語処理部のインタフェースを提供する。

ニュースリーダー部はグラフィカルなユーザインタフェースを提供する。ユーザには、以下の5つの要素からなる画面が提示される。

- リファレンスツリーリスト表示画面
- リファレンスツリー構造表示画面
- 類似記事検索ボタン
- ニュースグループ表示画面
- ニュース記事表示画面

プロトタイプモデルの動作例を図2に示す。

6. おわりに

討論型ニュースグループを対象とした知的ニュースリーダーの概要について述べた。ユーザへの情報提供法として、要約を表示するのではなく、ユーザのニーズに応じて適切な議論の流れを提示することにより、討論型グループの読者を適切に支援することが可能となる。

本システムはプロトタイプモデルが完成し、現在その試用と並行して、実用システムの開発を行っている。プロトタイプモデルの試用に際しては、心理実験およびテストセットの利用等を通して、客観的な評価が出来るよう心がけている。

現時点で、話題転換点の判定については、概ね満足できる結果が得られている。今後は、関連記事の収集の効率化、ユーザに提示すべき範囲の決定法の検討、ユーザへの提示方法等の検討を行い、公開可能なシステムへと改良を続けていく予定である。

謝辞

本研究の実施にあたっては、(株)東洋情報システムの松尾秀彦氏、高橋誠氏、藪田裕史氏、西俣憲幸氏、袋谷丈夫氏に、ひとかたならぬお世話になった、ここに記して深く感謝の意を表する。

参考文献

- [1] 小作浩美、井佐原均：話題の関連性に着目した知的ニュースリーダーの提案、平成7年電気関係学会関西支部連合大会(1995)
- [2] 井佐原均、小作浩美：Intelligent Network News Reader、1997 International Conference on Intelligent User Interfaces(1997)
- [3] 内元清貴、小作浩美、井佐原均：対話型ネットニュースグループにおける話題転換点の推定、言語処理学会第3回年次大会(1997)
- [4] 後藤康雄、林博之ほか：キーワード抽出自動システム、第37回システム制御情報学会研究発表講演会(1993)

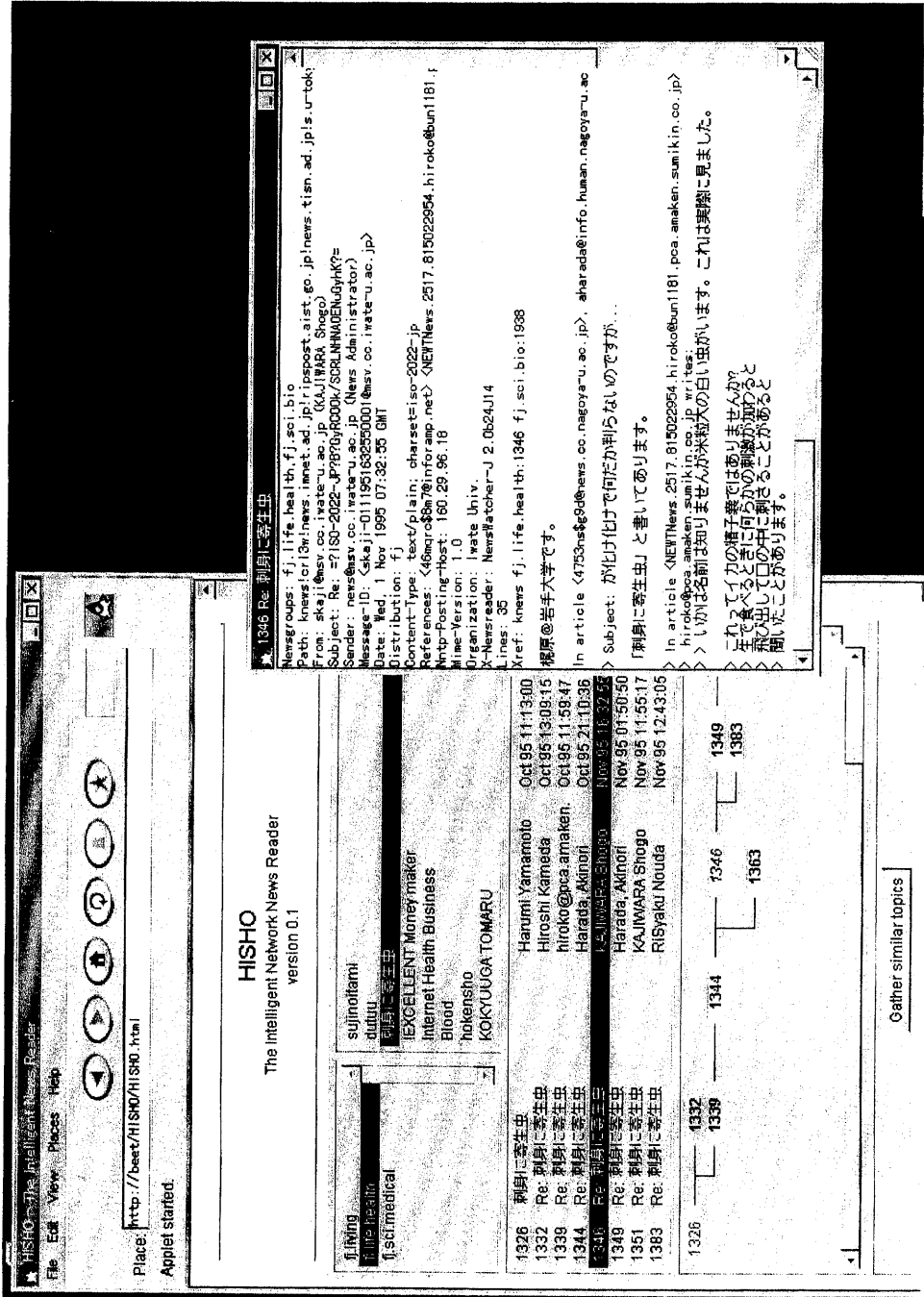


図2 知的ニュースリーダー・プロトタイプの実行例