知的検索のための単語細胞時空間成長腐敗モデル
－プライミング効果実現へのアプローチ－

1Aa－1

柴田正弘 伊藤英則

名古屋工業大学大学院 工学研究科 電気情報工学専攻

1. はじめに

近年，インターネットとコンピュータの普及により，ユーザが獲得できる電子化された情報は日々増加の一途とたどっている．多くの情報の中から必要な情報を即座に引き出す要求は今後ますます増えることが予想され，この要求に将来にわたって応えていくには，既存の情報検索技術に頼るだけでなく，新たな手法を開拓することが必要とされる．

検索技術における現在の問題点の一つは，検索結果を最適なデータのみを限定して出力できないということである．この問題を解決する方法として，概念の知識ベース（概念ベース）を使った類似性判別や観点の切り替えなどにより検索領域自体を切り替えることで検索結果を少数に限定して出力する手法が提案されている[5]．

しかしながら，概念によりクラスタリングされた辞書をベースとする検索システムは，概念ベース生成に使われるシソーラス（類義語辞書）に記載されている単語間の類似関係が固定的であるため，すべての人の検索要求に柔軟に対応できないことが予想される．

そこで本研究では，知的にしかも少数に限定して検索結果を出力するモデルとして，認知心理学で観察されている現象の一つであるプライミング効果を実現する単語細胞時空間成長腐敗モデルを提案する．

A model of Word-Cells growth in time and space for information search, Masahiro Shibata , Hidenori Itoh

Nagoya Institute of Technology, Dept of AI & Computer Science, Gokiso, Showa-ku, Nagoya 466, JAPAN

本報では，まず本提案モデルについて紹介し，モデルの基本要素の一つである「辞書内部時刻」の設定（初期化）手法が実験により妥当であるか評価する．

2. 提案モデル

2.1 本モデルの位置づけ

検索システムは，その検索目的により大きく2つに大別することができる．

一つは，辞書，百科事典，電子図書館あるいはネットワーク資源など，検索キーワードを使って，未知の情報を調べる目的でおこなう検索であり，もう一つは，ネットワーク経由で獲得した情報，読んだ情報，あるいはローカルで作成した文書など過去に扱った情報を再度引き出す目的での検索である．本研究は後者のためのモデルである．

後者の目的を実現する手法の一つにIPM (Incremental Path Method)と呼ばれる手法がある[4]．IPMは，観測データを順次ネットワーク状にデータベース（DB）配置し，すでに酷似データがDBに存在する場合は観測データとリンクされることで，検索時に類似データとして出力することを可能にする．

IPMと本提案モデルとの相違点は，IPMはDBを圧縮するという概念があるため，同じデータ，酷似データは同じDB空間に配置されるのに対し，本モデルでは必ずしも同じ時空間に配置されない点にある．

またこの配置のメカニズムこそがプライミング効果の実現を可能にする．

2.2 プライミング効果

心理学において，プライミング効果とは時間的に前に呈示された刺激語が後に呈示された刺激語の処理に影響を与える現象のことをいう[1]．工学的にいえば，この現象は，「データの保存時刻を糧として検索領域を動的に変化させ限定させることで，検索時間を減少させること」と言い換えることができる．本モデルでは，この検索領域の切り替えにDBへの入力データの観測時刻を用いる．

2.3 全体構成

本モデルにおける基本構成は，2次記憶・複数のインデキシングファイル・辞書内部時刻・検索エンジンの4つ部分からなる．

2.4 特徴

- データ保存過程において，生成時刻の異なる辞書どうしを連結する．(常識の生成)
- データは頻度や環境により辞書(IF)内部に想起出力可能な知識を形成する．(成長)
- ほとんどアクセスされないIFのデータへのリンクは徐々に圧縮され，消去される．(腐敗)

2.5 辞書内部時刻

音声認識分野において連続DPを用いて話題の要約が可能な技術がある[2][3]．

本研究では，この技術を用いて話題の区切れを抽出し，辞書内部時刻に設定することにした．(図1)

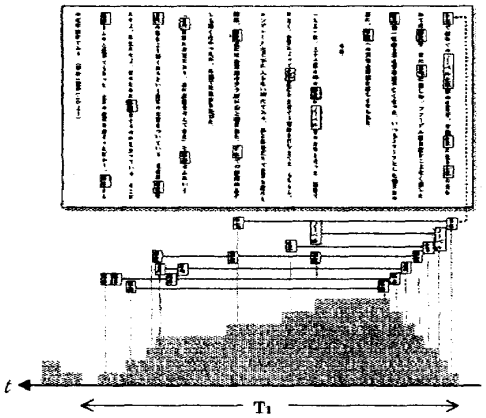

図1 時系列のクラスタリング

3. 実験

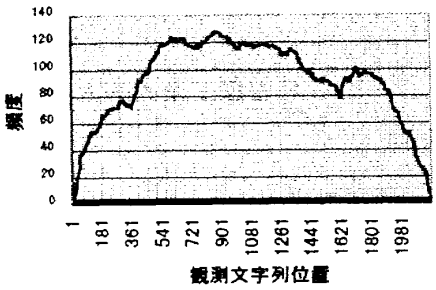

図2 新聞記事による時系列分割

コーパス…毎日新聞'96.1　形態素解析…ChaSen[6]

4. 考察

実験結果から新聞記事コーパスを使用しても，幾らかの山(谷)によりトピックごとに分割されていることを検証により確認した．

5. おわりに

本報では，本提案モデルについて紹介し，実験により，本モデルの特徴であるプライミング効果を実現するための時系列クラスタリングが適用可能であることを示した．

参考文献

[1] 斎藤洋典．第3章 心的辞書，岩波講座 言語の科学3 単語と辞書，pp.94-153，Dec.1997

[2] 岡 隆一．パターン情報の外形と言語情報の結合，人工知能学会誌，Vol.11 No.2 pp.9-10，1996

[3] 木山次郎，伊藤慶明，岡 隆一．Incremental Reference Interval-free 連続 DP を用いた任意話題音声の要約，信学技報，SP95-35，1995

[4] 遠藤 隆，中沢正幸，長屋茂喜，高橋裕信，岡隆一．音声と動画像の自己組織化ネットワークによるデータ表現とスポッティング相互検索，人工知能学会 SIG-J-9702-3 pp.15-20，Dec.1997

[5] 笠原要，松澤和光．概念の類似性における記号とパターンの統合，人工知能学会 SIGJ-9702-9 pp.51-56，Dec.1997

[6] 日本語形態素解析システム ChaSen http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html