

Wikipediaからの連想シソーラス構築プロジェクト

Our Association Thesaurus Construction Project from Wikipedia

伊藤雅弘^{1*}
Masahiro Ito¹

中山浩太郎²
Kotaro Nakayama²

原隆弘¹
Takahiro Hara¹

西尾章治郎¹
Shojiro Nishio¹

¹ 大阪大学大学院情報科学研究科

¹ Dept. of Multimedia Eng., Graduate School of Information Science and Techn., Osaka University

² 東京大学 知の構造化センター

² Center for Knowledge Structuring, The University of Tokyo

Abstract: Wikipedia, a huge scale Web based encyclopedia, attracts great attention as an invaluable corpus for knowledge extraction because it has various impressive characteristics such as a huge number of articles, live updates, a dense link structure, brief anchor texts and URL identification for concepts. We have already proved that we can use Wikipedia to construct a huge scale accurate association thesaurus. The association thesaurus we constructed covers almost 1.3 million concepts and its accuracy is proved in detailed experiments. In this paper, we introduce our project for constructing a high quality association thesaurus from Wikipedia

1 はじめに

近年、知識抽出のための有用なコーパスとして、ユーザ同士が協調してコンテンツを編集する Web 事典である Wikipedia に多大な注目が集まっている。Wikipedia は、Wiki を利用して構築された百科事典であり、文化、歴史、数学、科学、社会、テクノロジーなどの幅広い分野の話（記事）をカバーしている。Wikipedia には、2008 年 12 月の段階で約 270 万もの膨大な数の記事（英語のみ）が公開されており、市販の百科事典の記事数が数万から 10 万であることと比較してもその規模が膨大であることが分かる。

Wikipedia は様々な特徴を持っているが、中でも URL によって概念（記事）を一意に識別できその概念の詳細な説明があることと、膨大で密な内部リンクの存在が、概念どうしの関連度を定義した連想シソーラス構築に有効であることが筆者らのこれまでの研究で分かっている。具体的には、Wikipedia のリンク構造を n ホップまで解析することによって、概念間の関連を高い精度で求めることができる手法として pfbf を提案した。さらに、高いスケラビリティを実現する手法としてリンク共起性解析を提案した。これらの手法は、情報検索におけるクエリ拡張や語彙のあいまい性解消などの技術に用いられる。筆者らの手法に代表される、Wikipedia

のリンクを用いて概念間の関連度を計算する手法の大きな利点として、言語非依存な手法であるということがある。これらの手法を用いることによって、どの言語の Wikipedia であっても同様に解析可能である。

本論文では、筆者らの進める Wikipedia から大規模で高精度な連想シソーラスを構築するためのプロジェクトのこれまでの成果と今後の展望を紹介する。

本論文の以下では、第 2 章で関連研究について述べ、第 3 章で pfbf とリンク共起性解析について解説する。第 4 章では WikiSimi3000 Test Collection について述べる。最後に、第 5 章でまとめと今後の展望を述べる。

2 関連研究

Wikipedia では、Wiki によるコンテンツ管理を導入することにより、通常の Web コーパスや電子辞書とは異なる特徴を持つ。1 つ目は、ハイパーリンクによる記事どうしの参照である。各記事は、説明のテキスト、図表、そして別の記事に対する多数のリンクで構成される。従来の辞書や電子辞書では、機械可読なフォーマットで概念どうしの関係が表現されているものは少なく、概念どうしの関連を抽出するためには、説明文の中からさらに一度自然言語処理を行う必要があり、精度の低下を招く要因となっていた。しかし、Wikipedia の場合は Wiki をベースにしており、簡単に他の概念へのリンクを定義できることから、良質な概念どうしのリ

*連絡先：大阪大学大学院情報科学研究科
マルチメディア工学専攻西尾研究室
〒565-0871 大阪府吹田市山田丘 1-5
E-mail: ito.masahiro@ist.osaka-u.ac.jp

ンクが多いという特徴を持つ。

2つ目は、Wikipedia が高密度なリンク構造を持っていることである。筆者らは、予備実験として Wikipedia 内におけるリンク数をカウントしたところ、2006年9月の段階で約380万ページ（Redirect リンクを含む）に約8,000万の内部リンク（Wikipedia 内へのリンク）を抽出し、Wikipedia では閉じられた語彙空間の中で密なリンク構造を持っているということを確認している。

3つ目に、コンテンツの網羅性がある。従来、WWW を自然言語処理のコーパスとして利用する場合、その探索空間が膨大になりすぎることから、解析内容が収束しないもしくは偏ってしまうという問題があった。これに対し、Wikipedia は最新の幅広い分野の記事が網羅されており膨大な量のコンテンツが存在するものの、WWW の探索空間に比較するとそのリンク構造はサイト内で閉じられているため、現実的な時間での解析が可能である。

4つ目に、URL によって概念を一意に特定できるという特徴がある。自然言語処理においては、様々な局面で未知語の問題に突き当たる。たとえば、形態素解析において未知語が存在すると、解析の精度が大きく下がることは周知である。また「Apple」のような果物や社名など様々な意味が存在する多義語において、自然言語処理でそれぞれの意味を判別するには、前後の文脈で判断するなど高度な解析技術が必要であり、困難である。しかし、Wikipedia では URL によって一意に示される一つの記事（ページ）が一つの単語（概念）を表しており、多義を持つ単語には、意味に応じて別々の記事が用意されている。そのため、形態素解析における未知語の弊害もなく、多義語の判別も不要であり、高精度なシソーラス辞書の構築が可能である。

以上のような理由から、Wikipedia をシソーラス辞書構築のコーパスとすることは、多くのメリットがある。Wikipedia を解析してシソーラス辞書を構築する先行研究として、カテゴリリンクを利用した手法と TFIDF を用いた手法がある。以下にこれらの手法を解説する。

2.1 カテゴリリンクを利用した手法

カテゴリリンクは、Wikipedia における記事とカテゴリページ間の所属関係を表している。また、カテゴリページには他のカテゴリへのリンクによって、カテゴリ間の所属関係も定義されている。このカテゴリと記事の構造は単純な木構造ではなく、一つのページが複数の親カテゴリを持つことも可能であり、一部にはループも存在するネットワーク構造となっている。

Strube ら [9] は、WordNet のために提案されていた手法を Wikipedia のカテゴリ構造に適用する WikiRelate を提案している。WikiRelate では、二つの概念が与

えられた時に、それぞれの概念を表す記事が属するカテゴリの中でその二つの記事がどれほど近いかを基準に関連度を算出するという手法である。しかし、Gabrilovich らの研究 [2] で示されているとおり、カテゴリリンクを利用した手法は精度の点で問題がある。

2.2 TFIDF ベースの手法

TFIDF [7] は、Salton らによる文書中の重要なキーワードを抽出するための手法である。TFIDF は tf (Term Frequency) と idf (Inverse Document Frequency) の二つの指標を利用し、それらの積によって文書中の各語の重要度を計算する。 tf は文書中における特定の語の出現頻度であり、文書中に多く含まれる語が特徴語とされる。 idf は全文書中に、特定の語が出現する文書数の逆数であり、出現する文書数が多い語は idf の値が小さくなる。つまり、広く使われている一般的な語ほど特徴語としての重要度が低くなる。

この TFIDF を Wikipedia の各記事内のリンクに適用することによって、概念間の関連度を算出することができる。Wikipedia においては、一記事が一概念（語）に対応し、リンクは他の概念に対する意味的かつ明示的な関係を示す。そのため、TFIDF で記事内のリンクの重要度を計算することで概念どうしの関係性を抽出することができる。

この手法では、一つの概念の特徴ベクトルを抽出するには一つの記事に存在するリンク情報だけを解析すれば良いため、スケーラビリティは高い。しかし、それ故に記事の内容に信頼性がない場合やリンク数が少ない場合に、精度が低下する。

3 リンクの構造および共起性の解析

ここでは、筆者らの連想シソーラス構築プロジェクトにおいて、これまで行ってきた研究である $pfibf$ とリンク共起性解析について述べる。

3.1 $pfibf$

$pfibf$ [4, 5, 11] は、筆者らが提案するリンク構造解析手法であり、グラフ $G = V, E$ 内において n ホップ以内のノード同士の関係性を数値化することを目的としている。このとき、2記事間 (v_i, v_j) の関係の強さを計測する問題を考えた場合、関係の強さは以下の二つの要素に依存すると考えられる。

- 記事 v_i から記事 v_j へのパスの多さ

¹ $pfibf$ は、 $lfibf$ (Link Frequency - Inverse Backward link Frequency) から手法名が変更されたものである。

- 記事 v_i から記事 v_j への最短距離

つまり、記事 v_i から記事 v_j へのパスが多ければ多いほど（共通のリンク先や共通の参照元が多いほど）、記事間の関係性は強く、またそのパスの長さが短ければ短いほど強く関係すると考えられる。そこで、pfibfはpf(Path Frequency)とibf(Inverse Backward link Frequency)の二つの指標を利用し、それらの積によって関連度を算出する。pfは記事 v_i から v_j へのパスの多さと、各パスの長さによって決定され、全経路 $T = \{t_1, t_2, \dots, t_n\}$ によって次の式で表わされる。

$$pf(v_i, v_j) = \sum_{k=1}^n \frac{1}{d(|t_k|)} \quad (1)$$

ここで、 d は経路 t_k の経路長に応じて増加する関数であり、単調増加関数を利用する。

ibfは全記事中の記事 v_j が参照された数、つまり記事 v_j が持つBackwardリンク数の逆数である。この指標は、記事 v_j に対するリンクが多いほど小さい値になる。したがって、記事 v_i から記事 v_j への関連度はpfibfによって次の式で与えられる。

$$pfibf(v_i, v_j) = pf(v_i, v_j) \cdot ibf(v_j) \quad (2)$$

$$ibf(v_j) = \log \frac{N}{df(v_j)} \quad (3)$$

N は全記事数、 $df(v_j)$ は記事 v_j が持つ他の記事からのリンク数とする。つまり、pfibfに基づく記事 v_i から v_j への関連度は、 v_i から v_j へ多くの短いパスを持ち、 v_j のBackwardリンク数が少ない場合に高い値を示す。

3.2 リンク共起性解析

リンク共起性解析[3, 10]は、リンクの共起性を解析することによってリンク間(記事間)の関連度を算出する。リンクの共起とは、単語をリンクとして扱うということ以外、基本的な概念は単語の共起と同様である。つまり、リンクが共起するということは、特定の範囲において異なる二つのリンクが同時に出現することである。リンクの共起性解析では、リンクは参照先URLが同じなら同じリンクとみなされ、Wikipedia全体でのリンクの共起性を解析する。ここで、先に述べたように、Wikipediaにおけるリンクは、参照先の記事を一对一で表している。そのため、二つのリンクの関連度を求めることは、Wikipediaの記事が表す二つの概念の関連度を求めることと等価である。

ところで、Wikipediaを解析する時、同じ記事内での共起をカウントすると、リンク数の多い記事の場合、非常に膨大な共起の組み合わせが存在する。そこで解析範囲を近傍のリンクに限定するサイズ k のウィンド

ウを設定して、ウィンドウ内のリンクのみにおいてだけ共起しているとみなす[8]。このような指定されたサイズのウィンドウに基づいて計算された各記事のリンクペアの共起回数を、全記事で合算することによって、Wikipedia全体におけるそれぞれのリンクペアの共起回数を算出することができる。

共起回数から共起性を求める代表的な手法として、Cosine、相互情報量、Dice係数[6, 12]などがあるが、実験[3]によってCosineが最も高い精度を示している。以下にCosineの定義式を示す。

$$\text{Cosine}(x, y) = \frac{f_{xy}}{\sqrt{f_x f_y}} \quad (4)$$

ここで、 f_x, f_y はリンク x とリンク y がそれぞれ独立に出現する回数、 f_{xy} は x と y が同時に出現する、つまり共起する回数とする。

ここで、Hinrichらの研究[8]では、共起回数だけである組の関連度を算出する一次共起(first-order co-occurrence)に異論を唱え、ある組の語がどれくらい同じ語と共起しているかで関連度を算出する二次共起(second-order co-occurrence)を提案している。そこで、本研究でも二次共起によって共起性を求める。二次共起による共起性を求める際、各リンクにおいてどのようなリンクと共起するかという、各リンクの共起特性を表すリンクベクトルを生成する。リンクベクトルは、ベクトル空間モデルに基づく、リンクを次元、各リンクに対する重み(一次共起による共起性)を要素とする多次元ベクトルであり、リンク i の共起特性を表わすベクトル v_i は以下のように表される。

$$v_i = \{l_{i1}, l_{i2}, l_{i3}, \dots, l_{in}\} \quad (5)$$

ここで、 l_{ij} はリンク i, j 間の共起性である。このように作成されたリンクベクトルを、2ベクトル間のコサイン尺度によって、それぞれのリンクの共起性パターンがどれだけ同じかという関連度を求めることができる。

3.3 評価実験

本章では、各手法の評価のために行った実験について述べる。

3.3.1 実験概要

本実験の目的は、各手法によって構築された連想シソーラスの精度と構築時間を評価することである。実験対象の手法は、TFIDFベースの手法、pfibf、リンク共起性解析であり、それぞれの手法によってWikipediaから連想シソーラスを構築し、構築時間と精度を比較した。なお、pfibfにおいてはホップ数 n を2に設定

表 1: 連想シソーラス構築に要する時間

手法	計算時間 (秒)
TFIDF	278
pfibf	85,472
リンク共起性解析	220

表 2: 連想シソーラスの精度

手法	スピアマンの順位相関係数
TFIDF	0.57
pfibf	0.68
リンク共起性解析	0.65

し、リンク共起性解析ではウインドウサイズを 2 とし、Cosine によって共起性を計算した。

解析対象の Wikipedia のデータとしては、2006 年 9 月時点の英語版 Wikipedia のデータからノイズ記事を除去した、記事数約 82 万、総リンク数約 4,000 万のデータを用いた。ノイズ記事の定義は、トップページやカテゴリページなどの通常の記事ではないもの、記事内のリンク数が 5 つ以下のものである。

本実験では、データセットとして関連度計算手法の精度を計測するためによく利用 [2, 9] されている

「WordSimilarity-353 Test Collection」[1] を用い、「スピアマンの順位相関係数 (Spearman rank-order correlation coefficient)」によって正解集合との相関性を求め連想シソーラスの精度とした。

3.3.2 実験結果と考察

本節では、本実験の結果における計算時間と精度のそれぞれについて解説し考察する。

表 3.3.2 に、各手法の連想シソーラス構築に要する時間を示す。実験結果より、共起性解析と TFIDF の計算時間はほぼ同等であり、pfibf は 300 倍以上の計算時間を要することを確認した。pfibf は手法の特性上、 n ホップ先のリンクを再帰的に計算する。論文 [5] に述べられている近似手法を用いても、多量の計算が必要である。一方、共起性解析や TFIDF はリンク先を再帰的に処理することはしないため、少ない計算量に抑えられている。

各手法によって構築した連想シソーラスの精度を表 2 に示す。

まず、リンク共起性解析と TFIDF を比較すると、共起性解析は TFIDF より高い精度を実現している。これは、TFIDF では記事内に含まれるリンクのみを利用

し、記事 (概念) に対する特徴ベクトルを抽出しているのに対して、共起性解析では Wikipedia に存在するすべての記事を通して共起しているリンク組を抽出していることに起因する。各記事は限られたユーザによって編集されているので、各記事のリンク数や信頼性は均質でない。そのため、各記事のリンクによって得られる情報は必ずしも一般的というわけではなく、偏った内容となっている可能性がある。しかし、Wikipedia のすべての記事を通して出現する各記事へのリンクから得られる情報は、一部のユーザによる偏った情報ではなく、各記事に対する一般的な認識による情報となっている。つまり、ある記事へリンク付けを行うかどうかなどの、書き手によるリンク付けの偏りが存在した場合においても、共起性解析は様々な書き手を通しての統計的情報を得られるため、書き手によるリンク付けの偏りは平均化され、客観的情報が得られる。この理由により、TFIDF より共起性解析の方が高い精度を実現したものと考えられる。

また、リンク共起性解析と pfibf を比較すると、リンク共起性解析は pfibf の精度に迫っているものの、低い精度となっている。これは、記事間のリンク構造解析の方が高精度な関連度算出を実現できることを示している。

ここで、本実験によって得られた実験結果の精度が、構築した連想シソーラス全体においても有効であるかについて考察する。本実験で用いた 100 組の実験データセット中からランダムに数十組を抽出し、そのデータセットを用いて実験を 50 回行った。その結果、精度は多少変化するものの、提案手法と従来手法の精度の順位は 3 回しか変化せず、ほぼ同じ傾向を示した。このことより、構築した連想シソーラス全体においてもこのような傾向がみられるものと考えられる。

4 WikSimi3000 Test Collection

WordSimilarity-353 Test Collection は Wikipedia の関連度計算に関する研究 [2, 3, 9] において、もっともよく用いられているテストコレクションである。WordSimilarity-353 Test Collection は、353 組の単語を 13 人 ~ 16 人の被験者によって関連性を主観で 10 段階評価してもらい、その平均を関連度としている。しかしながら、このテストコレクションは Wikipedia 研究における関連度計算手法を評価する際のいくつかの問題を抱えている。まず一つ目に、単語ペアの数が非常に限られているということである。また、その単語も一般語に偏っており、Wikipedia のようなさまざまなドメインの概念を有するコーパスに対しては不十分であると考えられる。そして、さらに大きな問題として、定義されている単語の曖昧性が解消されておらず、

表 3: Wikisimi3000 Test Collection の一例

Concept1	Concept2	Relatedness
Apple Inc.	Steve Jobs	10
Apple Inc.	NeXT	6
Microsoft	Steve Ballmer	10
Microsoft	Solar panels	0
Personal computer	Laptop	9
Personal computer	Steel	0
Japan	Tokyo	9
Japan	University	0
Hong Kong	Kowloon	8
Hong Kong	Australia	3
Television	Broadcasting	10
Television	Greek language	0

Dog	Collie	8
Dog	Houston	1
Book	Magazine	9
Book	Scriptorium	2
Pasta	Noodles	9
Pasta	Food dye	5
Pokemon	Pikachu	10
Pokemon	Pop-culture	4
Police	Sheriff	9
Police	United States government	4
Halloween	Holiday	8
Halloween	Head	1

評価する際に Wikipedia の記事にマッピングするという作業を行わなければならない。しかし、このマッピングが正確であるという保証はない上、そもそもテストコレクション作成時に単一概念として被験者に提示されているわけではなく、単なる単語として提示されて実験されている点も問題である。

そこで筆者らは、英語版 Wikipedia の概念をベースにしたテストコレクション「WikiSimi3000 Test Collection」を構築している。構築手法は以下のとおりである。

1. Forward Link もしくは Backward Link の数が 5 以下の記事をノイズ記事として除外
2. リストページやカテゴリページなどの通常記事以外を除去
3. 残った 4000 万リンクを含む 82 万記事から、人物・地理・文化などのさまざまな概念から 72 記事をランダムに選択
(例: Malaysia, Television, Hong Kong, Book, Tea, United States, Coca-Cola, Cairo, Silk Road, Macintosh, Cheese, Bill Clinton, Japan, Olympic Games, Music, Halloween, Bicycle, Will Smith, Baseball, Elvis Presley)
4. 被験者が記事内から手動で関連リンクを選択
5. 被験者が各概念ペアの関連度を 0(完全に関係なし)-10(非常に強い関係) で評価

このようにして構築された本テストセットは、トータルで 3,836 の概念ペアを含む大規模なものとなっている。特筆すべきは、各ペアの関連度評価がそれぞれの

概念の記事を確認した上で行われるので、確実に単一概念どうしの関連度評価が行われているということである。構築したテストコレクションの例を表 3 に示す。

また先にも述べたように、筆者らの手法に代表される、Wikipedia のリンクを用いて概念間の関連度を計算する手法の大きな利点として、言語非依存な手法であるということがある。この利点を示す有効な手段は、多言語での評価を行うことである。しかし、当然のことながら WordSimilarity-353 Test Collection は英語のみの対応である。そこで現在、Wikipedia の概念をベースとした多言語テストコレクションを構築ために、日本語版 Wikipedia の概念をベースにしたテストコレクションも構築中である。現在までで、677 の概念ペアを、それぞれ 3 人の被験者が評価し、延べ 2031 の評価が集まっている。

5 まとめと今後の展望

本論文では、筆者らの推進する大規模な Web 事典である Wikipedia を解析し、連想シソーラスを構築するプロジェクトのこれまでの成果について述べた。まず、これまで概念間の関連度計算手法として pfibf とリンク共起性解析を提案してきた。pfibf は、Wikipedia の記事どうしのリンク構造を解析し、パス数とそのホップ数によって記事間の関連度を算出する手法である。またリンク共起性解析は、ある記事内にある他の記事へのリンクの共起性を解析することによって、リンク先記事が表す概念間の関連度を計算する手法である。評価実験の結果、精度は pfibf が最もよく、次いでリンク共起性解析で、最も精度が低い手法が TFIDF ベースのものであった。一方で、計算時間は pfibf が、他の二

つの手法に比べて非常に多くの計算時間を必要とすることがわかった。さらに本論文では、これまで使われてきたテストコレクションによる評価の問題点を指摘し、Wikipedia の概念をベースにした、関連度計算手法を評価するための新しいテストコレクションである WikiSimi3000 Test Collection を紹介した。

今後の展望として、概念間の関連度計算手法に関しては、SVM をベースとした手法を提案である。これまでの研究では、カテゴリやリンク構造、また共起性などの情報を個別に利用してきた。しかし、それぞれの手法には Wikipedia の記事の特徴の違いによって、得手不得手があると考えられる。そこで、Wikipedia の各特徴を WikiSimi3000 を教師データとして学習させることによって、精度向上を図る。また、WikiSimi3000 に関しては、さらなる概念ペアの数とその精度の向上を目指して、多くのネイティブの被験者によって評価を行っていく予定である。

謝辞

本研究の一部は、特定領域研究 (18049050)、科学研究費補助金基盤研究 (C)(20500093) およびマイクロソフト産学連携研究機構 CORE 連携研究プロジェクトの助成によるものである。ここに記して謝意を表す。

参考文献

- [1] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G. and Ruppin, E.: WordSimilarity-353 Test Collection (2002).
- [2] Gabrilovich, E. and Markovitch, S.: Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis., *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 1606–1611 (2007).
- [3] Ito, M., Nakayama, K., Hara, T. and Nishio, S.: Association Thesaurus Construction Methods based on Link Co-occurrence Analysis for Wikipedia., *Proceedings of Conference on Information and Knowledge Management*, pp. 817–826 (2008).
- [4] Nakayama, K., Hara, T. and Nishio, S.: A Thesaurus Construction Method from Large Scale Web Dictionaries., *Proceedings of IEEE International Conference on Advanced Information Networking and Applications*, pp. 932–939 (2007).
- [5] Nakayama, K., Hara, T. and Nishio, S.: Wikipedia Mining for An Association Web Thesaurus Construction., *Proceedings of IEEE International Conference on Web Information Systems Engineering*, pp. 322–334 (2007).
- [6] Peat, H. J. and Willett, P.: The Limitations of Term Co-occurrence Data for Query Expansion in Document Retrieval Systems., *Journal of the American Society for Information Science*, Vol. 42, No. 5, pp. 378–383 (1991).
- [7] Salton, G. and McGill, M.: *Introduction to Modern Information Retrieval*, McGraw-Hill Book Company (1984).
- [8] Schütze, H. and Pedersen, J. O.: A Cooccurrence-Based Thesaurus and Two Applications to Information Retrieval., *Information Processing and Management*, Vol. 33, No. 3, pp. 307–318 (1997).
- [9] Strube, M. and Simone Paolo Ponzetto: WikiRelate! Computing Semantic Relatedness Using Wikipedia, *Proceedings of National Conference on Artificial Intelligence and Innovative Applications of Artificial Intelligence Conference*, pp. 1419–1424 (2006).
- [10] 伊藤雅弘, 中山浩太郎, 原隆浩, 西尾章治郎: Wikipedia のリンク共起性解析によるシソーラス辞書構築, *情報処理学会論文誌:データベース*, Vol. 48, No. SIG20 (TOD 36), pp. 39–49 (2008).
- [11] 中山浩太郎, 原隆浩, 西尾章治郎: Web 事典からのシソーラス辞書構築手法, *情報処理学会論文誌:データベース*, Vol. 48, No. SIG11 (TOD 34), pp. 27–37 (2007).
- [12] 北村美穂子, 松本祐治: 対訳コーパスを利用した対訳表現の自動抽出, *情報処理学会論文誌*, Vol. 38, No. 4, pp. 727–736 (1997).