

WWWでの情報検索のためのナビゲーションプランニング

3AF-6

山田 誠二

東京工業大学 大学院総合理工学研究科

大澤 幸生

大阪大学 大学院基礎工学研究科

1 はじめに

インターネットは、研究者に留まらない広い領域のユーザを獲得し、急速に浸透しつつある。特にWWWにより、ユーザは、情報検索を容易に行うことが可能である。WWWではさまざまな情報が公開されており、ユーザは未知の概念を学習するときに、関連するWebページをアクセスすることで、WWWをあたかも百科辞典のように使うことが可能である。しかし、通信時間等のアクセスコストが大きい、適切なナビゲーションなしには、欲しい情報にたどり着くことが難しい等の重要な問題がある。

よって、ある概念を理解したい場合に、その概念を与えてやると、後はその概念を理解するために必要である関連したWebページを組織立てて、提示してくれるようなシステムの開発が、本研究の目的である。本研究では、“WWWでの情報検索を誘導すること”をナビゲーションと呼び、そのナビゲーションに利用可能な知識をナビゲーション知識、そして、その“ナビゲーション知識の自動生成”を人工知能におけるプランニングとして捉え、ナビゲーションプランニングと呼ぶ。本稿では、ナビゲーションプランニングの枠組の提案を行う。現時点で、システムは開発中である。

2 ナビゲーションプランニング

本研究では、ナビゲーション知識の生成を、プランニングとしてとらえる。なお、人工知能におけるプランニングとは、“行為主体がある目標を達成するために行為を実行するとき、その行為を目標達成に誘導するために使われる知識(プラン)を生成すること”を意味し、枠組と本研究での具体的要素(括弧内)は、以下ようになる。

- 行為：現在のWebページの内容を理解する。
- 状態：ユーザの知識状態。
- 初期状態：ユーザの初期の知識状態。
- 目標状態：ユーザが理解したい目標概念。
- オペレータ：U-Op (ユーザが、Webページを理解する。)
 - 条件：あるWebページを理解するために必要な条件知識 $C = \{c_1, \dots, c_i\}$ 。

- 効果：あるWebページを理解した結果、得られる効果知識 $E = \{e_1, \dots, e_j\}$ 。

3 KeyGraphによるオペレータ生成

ナビゲーションプランニングでは、WebページからオペレータU-Opの記述、つまり条件知識と効果知識を自動生成する。これらの知識は、対応する単語で表現され、その集合によりユーザの知識状態を記述する。条件知識とは、そのWebページを理解するために必要な知識なので、そのWebページ中出现する単語中に含まれていると考えるのが自然である。よって、Webページ中の単語から、どのように条件知識に対応する単語を抽出するかが問題となる。

まず、条件知識として考えられるのが、Webページ中で他のページにリンクが張られている単語である。そのような単語は、Webページの筆者により、明示的に条件知識として提示されていると考えられ、しかも辿る先のURLも明示されている。しかし、必ずしもすべての条件知識にリンクが張られているわけではないので、リンクの張られていない条件知識も抽出する必要がある。このような条件知識は、キーワード自動抽出システムKeyGraph[1]により自動抽出可能である。さらに、好都合なことに、KeyGraphでは、効果知識も同時に抽出できる。

KeyGraphにより、サブグラフ(概念の塊：隣接単語を枝で結んだときの一つの連結グラフ)を結び付けるキーワード群Bを効果知識の候補として、サブグラフ中の知識でかつキーワード群Aに枝で結ばれたキーワード群Bを条件知識の候補として取り出すことが可能である。

アンカー、KeyGraphは、それぞれ短所を長所を有する。即ち、アンカーは、重要な単語を正確に選ぶのに適しており、htmlのタグを基に抽出が容易である。しかし、ある単語が条件知識なのか、それとも別の意味で現在のページに関係するのかは断定できない。一方、KeyGraphで選択される単語は、正確にページの重要箇所を押さえているとは限らない。しかし、必ず候補を抽出でき、実験的に従来のキーワード抽出法より高精度である。そこで、両者を次の様に併用する。

1. KeyGraphでキーワード群A, Bを先に述べた様に抽出する。また、キーワードA(B)中の単語に対し、キーワード群B(A)と結び付く強さを重要度として与える。

2. キーワード群 A のうち、ページのタイトルタグ内の単語の重要度を上げる。
3. キーワード群 B のうち、アンカーとなっている単語の重要度を上げる。

即ち、*KeyGraph*によってキーワード群を多めに捕え、その正確さを高める目的のみにアンカーを用いるわけである。ここでステップ 2. は、条件知識の目安としてのアンカーに対して、効果知識の目安としてそのページのタイトルを用いることを意味している。ただし、やはりタイトルがついていない条件知識があるので、*KeyGraph*を補う為にのみタイトルを用いる。図 1 に *KeyGraph*で抽出されたオペレータの例を示す。効果知識（太線枠内）と条件キーワード（太下線）。著しく接近した単語間の単語も、キーワードとして取り出されている。

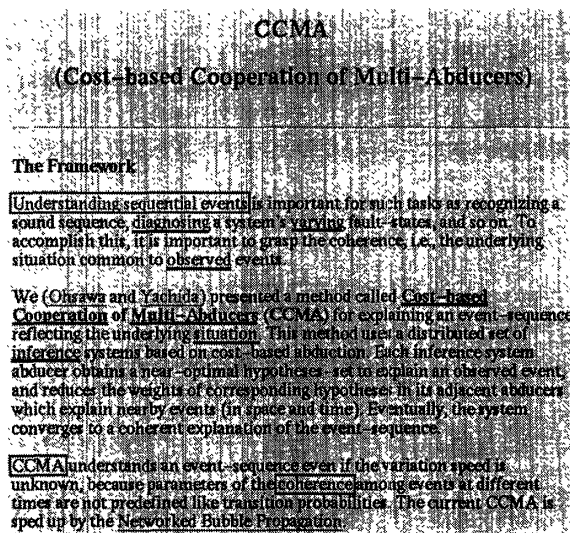


図 1 抽出されたオペレータ

4 ナビゲーションプランニング手続き

ここで、ナビゲーションプランニングの手続きを表す。基本的には、縦型探索の古典的プランニングに、前述のオペレータ生成を付加したものである。

各初期値：ゴールリスト $G = \{k_n\}$ ，状態リスト S ，オペレータリスト $O = []$ ，プラン $P = []$

1. G が、空であれば停止。 P が、求めるプラン。
2. G の先頭要素中のすべての知識が、 S に含まれている含まれている場合： G から先頭要素を取り除き、対応する U-Op(URL) を適用する。すなわち、U-Op(URL) を P の先頭へ追加、その効果知識を S に追加して、1. へ。
3. 含まれていない場合： G の先頭要素中の知識で、 S に含まれていない知識集合 D に対し、 D の部分集合を効果知識としてもつすべての U-Op(URL) を O から探す。

- (a) 見つければ、それらをオペレータ候補 C として、4. へ。
- (b) それ以外では、次のオペレータ生成を行う。
 - i. D の全ての知識を含む Web ページを検索エンジンで集める。
 - ii. その Web ページそれぞれについて、3の方法により条件知識と効果知識し、U-Op(URL) を生成する。
 - iii. 生成された U-Op(URL) に対し、評価関数 f により優先順位を決め、オペレータ候補 C とし、 O へ追加。

4. C 中の優先順位の高い U-Op(URL) を展開して、その条件知識 C' を G の先頭に追加して、1. へ。

なお、得られたプランは、オペレータの因果関係から、Web ページをどうたどればよいかを表す半順序関係であるプラングラフに変換される。

5 ユーザ知識の初期状態とオペレータの評価

次に、先の手続きにおけるユーザの知識の初期状態とオペレータの候補の評価関数 f を設定する。まず、ユーザの知識の初期状態は、最も一般的には、常識のようなものと考えられるが、常識を用意することは現実的に困難なので、本研究では、コーパスなどの情報を基にある程度領域に制限した基本概念を表す単語集合を用いて、ユーザの知識の初期状態を記述することを考えている。

また、オペレータの候補の評価関数 f は、追加される効果知識のサイズ等の一般に使われているヒューリスティクスに加えて、これも *KeyGraph* から得られる条件知識と効果知識のらしさの程度も用いて評価する。

6 ナビゲーションの提示

ナビゲーションプランニングで得られたプラングラフを、実際にユーザに提示する方法は、以下のものが考えられる。

- URL のリストを提示。
- グラフをそのまま提示。
- ハイパーテキストで提示。

7 まとめ

WWW での情報検索を誘導するナビゲーションの必要性を示し、そのナビゲーションに利用可能な知識の自動生成をプランニングとして捉えたナビゲーションプランニングを提案した。今後は、システムの実装とともに現実的な問題点を明確にしていく予定である。

参考文献

- [1] 大澤幸生, N. E. Benson, 谷内田正彦. 共起グラフを用いたキーワード抽出. 情報処理学会研究会報告 (96-FI-43), 第 6 巻, pp. 57-64, 1996.