

学位論文題名

最大エントロピー原理を用いた文献検索システムに関する研究

学位論文内容の要旨

検索者が希望する文献を、莫大な数の文献の中から効率的に検索するためには、文献に関する大量のデータを取り扱うための技法もさることながら、検索者が希望する文献をどのように表現し解釈するかの問題が重要なポイントとなる。つまり、検索者が文献検索システムに対して検索を依頼する場合に文献を指定する方法が問題であり、従来はブール代数による方法などが使用されてきた。

しかし、検索者の希望する有用な文献という概念は、人間が処理する微妙であいまいな情報と関連しておりブール代数だけでは表現しきれないという考えから、重み付きキーワードが考案され多くのシステムが採用してきた。この重み付きキーワードの方法は、文献をベクトル空間の一つの点としてとらえることにより検索方法を発展させてきた。しかし、重み付きキーワードの“重み”はそのキーワードの重要度を表わす値であると解釈され、この値を計算する方法によって結果が異なるために計算の意味づけが問題になる。そこで Cooper は“重み”を条件付き確率の値であるとし、統計的な計算を用いて有用性の高い文献を出力する方法を提案した。この方法は、“重み”などの意味づけが数学的に明確であるうえ、最大エントロピー原理により検索結果の妥当性が保証されていることや、多くの有益な性質のために広く支持されている。

本研究では、この Cooper の提案した最大エントロピー原理を用いた文献検索における幾つかの問題点を提示し、それを解決するために Bayesian network、ファジー理論などを用いて新しい文献検索法を提案する。

第一章では、本研究の背景と目的について述べる。

第二章では Cooper の最大エントロピー原理を用いた文献検索法について説明する。確率的検索法で広く用いられる“確率ランキング原理”について述べ、コロモゴロフの確率の公理を文献集合上にあてはめることにより、最大エントロピー原理を用いた文献検索で重要な“有用性の確率”を定める。そして、この検索法の特徴と問題点を挙げる。

第三章では、近年エキスパートシステムや確率的推論の研究において注目を集めている Bayesian Network と呼ばれる確率モデルをについて説明する。これは、有向グラフを用いて確率変数とその間の確率的関係を表現しており、ネットワーク全体として確率モデルになっている。その外見や働きはニューラルネットワークのボルツマンマシンと多くの共通点を持っているが、その発想は異なることからきており、理論と応用の両面において異なった性質をそなえている。

第四章では、第三章で述べた Bayesian network を用いて最大エントロピー原理を用いた文献検索法を実現するための具体的な方法を述べる。本研究では、Bayesian network の確率的推論法を用いて、“有用性の確率”の計算を行なう確率推論システムを実現している。この確率推論システムは、キーワード同士の確率的関係の情報を文献データから学習することにより記憶、保持するためのネットワークと、有用性の確率の推定のための節点からなる。これにより、最大エントロピー原理を用いた文献検索法では、キーワード同士の確率的関係をどのように決定するか不明確であるという問題を克服している。さらに、ネットワークの自己組織化を利用することによって検索者が入力したキーワードに関連するキーワードを選出する機能が実現され、再検索時における検索者の負担を軽減することが可能になる。そして、ネットワーク構造の改良や拡張により、推論システムの調整や改善な

どの興味深い応用の可能性を開いた。

第五章では、第四章で提案した検索法を実現する上で必要となる制約条件付き最適化問題の解法について述べる。この制約条件付き非線型計画問題では、制約条件が満される可能領域はパラメータの許容領域の一部であり検索条件に対応して変化するために、最大勾配法を用いた探索における出発点を求める方法や、探索の過程で検索条件を満す領域からはずれてしまわないようにするための工夫が必要である。これを実現するために検索条件の直交化の方法を提案する。

第六章では、我々の提案した文献検索法を実際に計算機上を実現し、キーワード間の確率的相関関係を表わすキーワードネットワークを、文献データから自動的に作成する実験、実際に検索リクエストを入力し検索結果を求める実験、第四章で提案した関連キーワードの選出機能の再現率、適合率を求める実験を行なった。また、作成した文献検索システムのユーザインターフェイスについても説明する。

第七章では、Cooperの提案する文献検索法において、ユーザが入力する検索リクエストから計算される確率的な条件が厳しすぎて有用性の確率が計算できなくなるという問題を解決するために、従来の検索条件をファジー集合論を用いて拡張し、最大エントロピー原理を用いた文献検索法でファジー検索条件を取り扱うことができるように改良する方法を述べる。このファジー化された検索法を用いることによって、検索条件を厳密に満す検索結果が無かったとしても、検索条件に最も近い検索結果を出力することができる。その結果どのような検索リクエストに対しても必ず検索結果が得られ、検索条件をより柔軟に指定することが可能になる。

第八章では、本論文の結果をまとめ我々の提案した文献検索法の特徴、今後の課題を示す。文献検索システムの研究においては、検索者が思い通りの検索結果を得るためには、検索者自身が試行錯誤的に検索リクエストを修正する作業は不可避であり、かつ最も手間を要する。次の検索リクエストに使用するキーワードの候補を表示する機能、つぎにどのような検索リクエストを行なうかの手掛りとなる検索結果が必ず出力されるという機能、検索条件に強弱を付け検索結果を細かく調整する機能は、一度の検索課程で思い通りの検索結果が得られないような状況で検索の利便性を向上し、結果的に検索者の負担を軽減すると期待できる。

学位論文審査の要旨

主 査 教 授 伊 達 惇
副 査 教 授 新 保 勝
副 査 教 授 佐 藤 義 治
副 査 教 授 宮 腰 政 明

学 位 論 文 題 名

最大エントロピー原理を用いた文献検索システムに関する研究

インターネットの普及や情報処理機器の進歩に伴って、利便性の高い情報検索システムが要求されるようになり、情報を効率的に検索するための方策として、検索者の要求を的確に表現するための研究が多数行われている。従来から、キーワードを用いたブール代数検索方式、これを改良した重み付きキーワード方式、さらに重み付きキーワードの“重み”の計算法に任意性がある点を改良した統計的な検索方式などの研究が進められ実用化も進められてきた。統計的な検索方式の一つである最大エントロピー原理を用いた検索方式は、検索結果の妥当性が保証されていることや検索リクエストに高い表現力があるなどの有益な性質のために広く使用されている。

しかし、最大エントロピー原理を用いた既存の文献検索方式においては、検索のプロセスで使用するキーワード間の確率的相関関係データの作成法が定められておらず、またごく稀に、検索リクエストとキーワード同士の確率的関係のデータの組合せによって検索結果が計算できない状態になるという問題がある。

本論文では、ベイズネットワークと呼ばれる確率的ネットワークモデルの学習機能を用いて、キーワード間の確率的関係のデータを自動的に作成する方法を考案し、同時に関連するキーワードを選出する機能も実現している。次に、情報検索システムにファジィ情報処理を導入することによって、例外なく全ての検索要求に対して検索結果を出力することを可能にし、合せて検索条件の強度を指定する機能を付加することによって、より柔軟な検索を可能にしている。

第1章では、本研究の背景および、目的、本論文の概要について述べている。

第2章では、最大エントロピー原理を用いた文献検索法について述べている。確率的検索法において広く用いられる確率ランキング原理および、有用性の確率について述べ、さらにこの検索法の特徴と問題点を挙げている。

第3章では、エキスパートシステムや確率的推論の研究において、近年注目を集めているベイズネットワークについて説明している。これは、有向グラフを用いて確率変数とその間の確率的関係を表現する確率モデルであり、学習機能あるいは確

率的推論を行う機能を有するものである。

第4章では、有用性の確率の計算を行う確率推論システムをベイズネットワークを用いて実現するという著者が開発した手法について述べている。さらにこれを用いて、キーワード同士の確率的関係のデータを文献データから自動的に作成する方法を提案している。次に、ネットワークの自己組織化を利用することによって検索者が入力したキーワードに関連するキーワードを選出する機能を実現し、再検索時における検索者の負担を軽減する方法を述べている。

第5章では、最大エントロピー原理を満足する確率分布をベイズネットワークを用いて計算するための、制約条件付き計画問題の解法について述べている。連立一次方程式から成る制約条件を直交化し、これを最大勾配法に適用して効率を向上させている。

第6章では、提案した文献検索法を実際に計算機上に実現し、文献データからキーワード間の確率的相関関係を表わすキーワードネットワークを自動的に作成するシミュレーション実験、実際に検索リクエストを入力し検索結果を求めるシミュレーション実験を行い、第4章で提案した関連キーワードの選出機能の再現率、適合率を求めている。

第7章では、最大エントロピー原理を用いた既存の文献検索法において、稀に有用性の確率が計算不可能になることがあるという問題を解決するために、ファジィ集合論を用いて従来の検索条件を拡張し、ファジィ検索条件を取り扱うことができるように最大エントロピー原理を用いた文献検索法を改良する方法を述べている。この方法を用いることによって、どのような検索リクエストに対しても必ず検索結果が得られるようになり、さらに検索条件の強度を指定する機能を付加することによって検索の柔軟性を増大させている。

第8章では、本論文の結果をまとめ提案した文献検索法の特徴、今後の課題を述べている。

これを要するに、著者はベイズネットワークおよびファジィ理論を用いることによって最大エントロピー原理を用いた文献検索法を開発し、検索の利便性、検索結果の妥当性およびユーザへの個別対応の柔軟性を向上させるなど、情報検索の研究において新知見を得ており、情報工学の発展に対して貢献するところ大なるものがある。

よって著者は、北海道大学博士(工学)の学位を授与される資格あるものと認める。