

(2) 自然言語処理システムの概要

(発表者: ICOT研究所 第4研究室長 内田 俊一)

質問: 文脈文法に関してお触れになったと思いますが、論文ではここで具体的な提案を行っていらっしゃるのですか。

コンテクチャルな表現、文脈情報を文法に入れる特別な提案を持っていらっしゃるのですか。

回答: いいえ。私のご覧にいらしたスライドは、自然言語理解システムの理想形を示した訳であります。特に上の半分がそうです。

質問: それでは、文脈文法というようなものは可能ですか。

考えられていますか。

回答: はい。我々はそれを今考えているところです。と言いますのは、我々は非常に複雑な形態的、また構文的な解析をしなければならないということで、もしより良い方法で、この文脈文法など入れていく方法があったとしたならば、もちろんこれは良い方法になります。

しかしながら、そのような方法に余り依存しすぎると言うことは、好ましくないと思っております。その前に我々はまずいろいろな実験をするべきだと思っております。と言いますのは、どの部分を文法に入れるべきか、そしてどの部分を辞書の方に持っていくべきかということとは難しいことでもあります。

質問: 私が感じておりますのは、文脈というのは、有限なコンセプトであると思いますが、文生成におきましては、常に関係してくる訳でありますけれども、コンテキストというのは、用途に依存するものでありますか？

回答: 例えばサンプル・センテンスを選択す

る場合に、例えば日本語のテキストから選びますね、これは非常に難しい選択になる訳です。私が例えば物理の教科書からサンプルを採ったとしたならば、ショートカットが出来る訳ですね。コンセプト・ディクショナリーとか、シンタクティック・インフォメーションでやれる訳です。ですから確かにアプリケーションにかなり依存してくると思います。

現在はより難しい複雑な問題を定義しようとしているのだと思います。

質問: ニコラス・オスラです。イギリスからです。

今回の発表と、その前の方の発表とは共通点があると思います。例えば、文脈情報ということを考えますときに、これは例えば新しい、それに対するレプレゼンテーションとか、新しいルールを考えなければならない。決してそれを非常に多くの知識を集積すればいいというものではございません。解釈する為にですね。

従いまして、前の方がおっしゃいましたように、例えばコモンセンス・ノーレッジというようなものが考えられる。それを非常に強力なルールで適用して、常識・知識というものを表現していく。それは、ただただヒューリスティックにいろいろな常識・規則というものを集積すれば出来上がるものではございません。

特に今朝触れられた点であります。サイモンさんがおっしゃった点でありますけれども、他の計画で今の仕事を補助するようなものがございませぬか。例えば大規模な概念辞書、またはコモンセンスの集めたようなもの。そのよう

なもの、例えばルールベースのアプローチのものが問題を解決するというふうに思ってもらえますか。と言いますのは、これは少数意見だと思うのですけれども、そういった意見はどう考えていらっしゃるのでしょうか。

回答：そうですね、私共の立場と言いますのは、ちょっと最初の第1研究室のしているものとは違う訳であります。

私の個人的な見解になりますけれども、多くの提案が枠組み、知識表現の枠組みまたは問題ソルバーの解決に対しますフレームワークが提案されております。この例えばいわゆるコンセンサスを記述する方法として、フレームワークが提案される訳でありますけれども、しかし実際には、評価する方法がない訳ですね、例えばある程度の量の文を使っての評価というものが出来ていない。従いまして、現在我々はある実験のプラットフォームを作ろうとしている訳です。そしていろいろな仮説を評価したい訳です。いろいろなコンテクスチャル・ノーレッジを集積して、例えば先程の質問の方がおっしゃいましたように、コンテクスト・グラマーとか辞書が適用できるか否かを評価したい。

そして、このようなグラマーとか辞書というものは、現在、現存しない訳でありますから、従いまして我々は実験を繰り返して行って、そしてその集積結果としてやってみたいということです。

質問：と言いますと、あなたのストラテジーと致しましては、スタートとしてはまずテキスト解析からですか、これがまだやられていないものに適用していくということですか。そして、全般的にまず解析に対する全般的な自然言語解析に対する知識を集積して、それからプリンシプルを引き出してこうということですか。

それでは、第2の質問の部分ですけれども、パラレリズム、並列化でありますけれども。

回答：ちょっと、お答えさせて下さい。

そうですね。今おっしゃったことはその通りです。しかし、我々は現段階にやっと到達したわけです。我々は実験出来るという環境が与えられた。すなわち、コンテクスチャル解析というのが、非常に強力なローレベルの処理、プロセッサがなければハイレベルなプロセッサと言いますのは、全く意味がなくなってしまう訳であります。現在は非常に強力なライブラリーが、ランゲージ・ツールボックスの中に入れておられます。従いまして、ここで我々の実験が上部の評価ということに対しても有意義になってくる訳であります。

我々は同じく、また例えば状況理論とか、そういった仮説も研究している訳であります。従いまして主に状況理論を評価して、そしてこのようなシチュエーション・セオリーに基づく知識評価ができるかどうかとも研究したいわけです。

質問：それでは、自然言語の作業を例えばスピーチ処理に統合していくというようなことがありますか。

と言いますのは、これが並列化に伴う主なアプリケーションとして出てくると思います。並列、例えば並列処理によって出てくる、これは一般的には、必要、いまおっしゃったような処理に関しては必要ないというふうに考えられているのですが、どうでしょうか。

回答：その前に音声認識ですか、その前に私の個人的な見解ですけれども、現在我々が専念しておりますのは、音声キャラクター・ストリームに変換されてからあとの処理するということあります。従いまして、音声認識というようなことをスタートすることはないと思います。高レベルの処理は逐次的なので、強力な並列コンピューターシステムのようなものは、必要ないというふうにおっしゃいましたけれども、私は反対です。

と言いますのは、ごく全般的に申しまして、もし人間がより勤勉に仕事をしたくないという

のでしたら、マシンがサポートするべきでありまして、人間にとって代わるべきです。そうしますと何らかの補完するものが必要になってくる訳であります。例えばロジック・プログラミング・エンジンのインプリメンテーションということについて言えば、人間のプログラマが

レージーカバーしたくないというのでしたら、マシンにカバーさせなければならないことになります。

このように、常に強力なコンピュータは必要になると思います。