

随筆：「認知科学関連書を読んで」

春になって、だるい日が続きます。ですが、未夢も大体形ができてきて、これから本格的に意味理解システムとして完成させていきたく思います。そこで、参考書として次の本を読みました。この5年間に考えたことがつまびらかに記述されているのを見て、先人の偉大さを改めて感じました。と同時にこれから意味理解に本格的に向き合うための技術が見えてきた感じでうれしく思いました。ブレークスルーしていく自信がついた気がします。

#### 【参考文献】

- ・心の計算理論（コレクション認知科学） 往住彰文著 東京大学出版会
- ・視点（コレクション認知科学） 宮崎清孝、上野直樹著 東京大学出版会
- ・比喩と理解（コレクション認知科学） 山梨正明著 東京大学出版会
- ・理解とは何か（コレクション認知科学） 佐伯胖著 東京大学出版会
- ・脳型コンピュータの実現に向けて 甘利俊一編著 サイエンス社
- ・PDP モデル D.E.ランメルハート著 産業図書

#### 1. 弛緩法とコネクショニズム（ニューロコンピュータ）

未夢も曖昧性管理のために、本格的に解析結果の評価をしていく段階になりました。無数にある解析結果から最適と思えるものを選択していくには、文章の中の多数の文の解析結果をつきあわせていかねばなりません。それは、全体の文脈が決まらねば部分の文の意味が決まらない、反対に部分の文の意味が決まらねば全体の文脈が決まらないという、ジレンマの問題に遭遇することです。この解決には弛緩法によるしかないでしょう。弛緩法とは、選択項目にプラスのプライオリティを持たせ、またマイナスのプライオリティを持たせ、互いに評価しあい、プラスのプライオリティのうち、もっとも高いプライオリティの項目を選択することで、最終の最適解をあぶり出していくというアルゴリズムです。解析項目間にプライオリティを定めて、連想関係を持つこととなります。

一方で、曖昧性の項目は無数にあるため、組み合わせると組み合わせ爆発を起こします。このため、未夢では項目を最低の粒（単語）において、粒の間の関係はポイントで表現し、文とか、文章とか単語の連鎖も単語を組み合わせないで表現しています。もちろん、曖昧性項目にはプライオリティを設けて弛緩法に対応させています。

この技術はコネクショニズムと同じ表現ですね。ニューロコンピュータのノードが単語で、パスが単語連鎖のポイントで、パスの重みがプライオリティというわけです。とするならば、ニューロコンピュータで培われている技術はそっくり曖昧性管理、自然言語処理技術として流用できるということです。特に、弛緩法としてボルツマンマシンを採用することは魅力的です。

大体が、曖昧性評価とか文脈決定とかは学習機構が必要で、状況が無数にあるため人手によるプログラミングには対応できないことだと思えます。確かに今のニューロコンピュータとはかなり違ったものですが、新しいニューロコンピュータのモデルとして捉えていける形態的な類似性があります。今の未夢はコネクショニズムな格好になっているので、これを理論化するにはニューロコンピュータの新規の形態として捉えることは有用だ

と思うのです。弛緩法が見通しよくなりますから。

## 2．曖昧性、多義語選択評価関数（アルゴリズム）と PDP 調和理論

本「PDP モデル」の調和理論は圧巻でしたね。パターンをプリミティブ（本では「アトム」といっている）なパターン要素のコネクション（パスと重み）で表現し、弛緩法を展開し学習システムを実現、パターンを選別するネットワークを構成していくというものです。このプリミティブをベースにした、コネクションで任意のパターンを表現していこうというのがいいですね。未夢の評価システムに使えと感しました。

多義語の選択はもっとも重要な曖昧性処理です。意味要素をプリミティブを見つけて、それで各単語を表現していくことになります。多義語の選択は、このプリミティブのコネクションを学習によって変化させて最適解にもっていけるように機構を作り込むことによって行います。意味要素のプリミティブは単語のプリミティブ記号だけでなく、句とか節、文脈の要素全般です。数学的にやったらものはできなくなると思えるので（要素が多すぎて）、学習アルゴリズムを工夫しなくてはいけないと思いますが。柔軟にプリミティブを追加できるようにするとか、記号処理を学習によって成長させることを考慮して設計することです。

ノード自律拡張型のボルツマンマシンですが、この場合、背景として文章全体の状況（文脈）があって、学習するノードは文とか単語ですね。多義語、多義句、多義文なんてあって、時々刻々選択していく必要があります。へんなニューロコンピュータになるかもしれません。

## 3．パターン発見と学習機構

学習機構を実現するには、パターン発見能力は必須で持たねばなりません。このパターン発見能力も PDP 調和理論を応用して実現できる気がしています。コネクションの誤差もコーパスとしてスタックしていき、あるコネクションインスタンスが別の評価を必要とされたときに新しいカテゴリー創生させるようにするのです。基本的なカテゴリーシステムはこれで作れるでしょう。

確かに、推論とか現実世界を分析することによりカテゴリーが生まれることもありましょう。しかし、その場合も基盤機構として、PDP 調和理論の応用的カテゴリー創生機構を持つことは有用でありましょう。かなり PDP 調和理論に還元できると思うのです。

## 4．これからの未夢拡張計画

永続化と耐久テストがクリアできた今、未夢をより包括的な、本格的な意味理解システムへと成長させるつもりです。今は対症療法で作成した未夢なので、ちょっとラビリンスみたいなところがあります。できるだけ早い時期にスッキリした整頓された構造に直していきたいと思っています。

基本的に次の項目を予定しています。

- (1) 比喩（擬人化、擬物化など）への対応
- (2) 各種構文サポート（構文解析結果の類型化）
- (3) 記号の接地（オントロジーベースにする）

- ( 4 ) 多義語処理の充実
- ( 5 ) 鍵と鍵穴システムの実現
- ( 6 ) 評価系の充実 ( 弛緩法の導入 )
- ( 7 ) 学習システムの導入
- ( 8 ) フレームシステムの導入

#### 4 . 1 比喩への対応

擬人化、擬物化機能をサポートしていかなばなりません。単語の意味記号を書き換えて使用していく必要があるのです。

( 例文 ) 私は博士だ。 ---> 「博士」の意味記号は+role です。これで有効文になります。

( 例文 ) 私はカエルだ。 ---> 「カエル」の意味記号は+life です。人間がカエルというのは非分です。ですが、擬人化しているときには「カエル」が「人間」として処理されねばなりません。そこで、意味記号を+human に書き換えて意味理解処理をしていく必要があるのです。これは、ベースの自然言語処理の機能を厳密な解析をしていくシステムとして作る必要があるので、このような特別な処理を必要とされるからのことです。

擬人化が行われているかどうかは、単語や句の共起関係から判断していく必要があります。面倒ですがやっていくしかありません。文脈情報に擬人化情報を持たせねばいけません。

比喩はよくある参照技術だと思うのです。「信子」の単語を登録するのに「信子 = 明雄」とする。すると「明雄」で定義されている意味記号群全部を一度に「信子」に設定できるわけです。この参照技術を大々的に利用していくと比喩システムになるのではないのでしょうか。

( 例文 ) 信子は花だ。 ----> 「花」を擬人化して新しい単語を作り、「花」から意味記号を引き継ぐ。

( 例文 ) 連休には三島が来る。 ---> 「三島」という地名を擬人化して、「三島」から意味記号を引き継ぐ。当然、三島に住む住人へのポイントも引き継ぎ、「三島」とはこの人たちであると推論して処理していく。

( 例文 ) そんなの泥棒が来て縄をなうようなものだ。 ---> 文の拡張意味要素に「泥棒が来て縄をなう」の本来の意味「間に合わない」を添付することになる。

#### 4 . 2 各種構文サポート

( 1 ) 名詞構文には格表現したシャドウな解析結果をアンカーとなる名詞項目に追加できるようにする。

( 2 ) 感情表現は次のような表現でシャドウとして動詞項目に持つ。

・ I feel that( . . . . . )。

・ I think that( . . . . . )。

( 3 ) 使役表現は次のよう表現でシャドウとして動詞項目に持つ。

・ I make that( . . . . . )。

#### 4.3 記号の接地

脳の基本機能は、

- (1) 好感・・・ドーパミン
- (2) 苦痛・・・ノルアドレナリン
- (3) セロトニン・・・調整

であるといえます。基本的な接地は好悪の2つであると思いませんか。

視床、視床下部があり、大脳辺縁系、旧皮質があり新皮質がある。また脳幹もある。そして、事細かに機能領野がある。言語野はその一つとして現れています。特にミラー細胞系があって、接地問題を解決しているように思えます。

#### 4.4 多義語処理

分かち書き化と係り受け解析（構文解析）のところでは、多義語という曖昧性を解決していかなばなりません。基本的に鍵と鍵穴システムとして作り込んでいくことになります。

#### 4.5 鍵と鍵穴システム

曖昧性には全て鍵穴を設定します。基本的に単語の意味記号として定義します。鍵は単語列パターン（コーパス）に設定し、特別なプロセスで鍵穴と鍵のマッチング、項目の選択を行います。選択は項目のプライオリティを高めることで行います。また、鍵穴はプラスとマイナスの両方があるべきです。

##### (1) 鍵穴の定義

+keyhole@カテゴリ記号@バリュー----->プラス鍵穴

-keyhole@カテゴリ記号@バリュー----->マイナス鍵穴

##### (2) 鍵の定義

+key@カテゴリ記号@バリュー

#### 4.6 評価システムの充実

Google はインターネットのホームページの評価アルゴリズムによって、のし上がってきたといえます。それぐらいに評価システムとは重要な意味を持っています。洗練されたアルゴリズムを得ることは知的作業です。

今は基本的にコーパスとのマッチングを取って、マッチ数の多いものを高いプライオリティとしています。コーパスにも重要なコーパスと連想を持つことで重要性を増していきます。そんなこんなで、評価システムも学習によって最適化を図れるようになっていくべきでしょう。学習システムの構築も重要になります。

さらにここには次のような機構の実現が必要になってくるでしょう。

- (1) 文章、会話内容評価
- (2) 情報の重要度評価
- (3) 問いを発するためのプランニング
- (4) 話者のプロフィール獲得
- (5) 話題への思い入れ、話者の評価、話者関の関係評価
- (6) 自分の性格と話者との関わり方の評価

また、メタ評価システムによって、評価システムを学習によって充実させていく必要があります。そういうことで、評価システムの要素にも意味記号が張られて管理していく必要があります。

#### 4.7 学習システム

PDP 調和理論をベースに強化学習で実現していくつもりです。PDP 調和理論については「PDP モデル」に詳細な説明が載っています。簡単に説明しますと、一般の記号とかパターンとかを、オントロジー的プリミティブ記号への連想で表現するものです。連想は重みです。基本的にニューロコンピュータのボルツマンマシンとして表現します。

記号 = 重み  $i$  \* プリミティブ記号ノード

プリミティブ記号ノードは (プリミティブ  $a$ , プリミティブ  $b$ ,  $a$  と  $b$  の連想) だと私は思っています。

オントロジーをしっかりと作り込むことが先ずやらねばならないことですね。コーパスのスケルトンとか、単語のスケルトンとを持っていて、それを巧く利用して自律学習、カテゴリー創生ができるように死体と思っています。人手と機械学習の調和を図れるようにする。

学習システムはカテゴリーの獲得です。カテゴリーは入力と出力の関係をブラックボックス化し、記号として扱うシステムです。つまり、カテゴリー発見にはいかなる出力項目にパターンが関係するかということを解析していくことです。学習機構は評価機構との関係が深いのです。

#### 4.7 フレームシステムの導入

意味を理解するには、単語に定義されている情報だけを組み合わせしていくのでは限界があります。そこには、話題を推測するためのフレームが必要になってくるのです。文から得られた情報を埋め込み全体を見据えていくフレームをベースに文章を解析していく必要があります。なぞなぞ文みたいなものでは、話題の枠組みが分からないと理解できないものがあります。フレームとは文脈と等価なデータベースであるといえます。

フレームは動画システムでなくてはなりません。そして、文章を解析しながらどの動画システムについて語っているか投票していく必要があります。また、動画システムの永続化は Prolog 文で記述してますので、未夢をベースにするならば、フレームの表現は Prolog にするつもりです。

また、意味理解には各登場人物 (アクター) の立場、文脈内の位置を精確に把握していかなければなりません。登場人物毎にフレームを設けて、必要に応じてそのフレームを解釈し直して、会話ができるようになっていなくてはなりません。ミラーニューロン機構ですね。特に、まねをしたり、他者の行動を自分の行動パターンに対照して理解していかなければなり

ません。フレームも Prolog 文セットして情報を保持することになると思います。

## 5 . おわりに

研究とはなかなか進まないものだ。こうすればいいのにとっても、なかなか思うようにはプログラミングができない。やっとならしてみれば、理想型とはほど遠いプログラム構造になっている。でも、ストラグルするうち少しずつ理想型に近づきつつあると思えることもあったりする。何度もコーディングし直し、プログラムの構造の理想型をその向こうに見えるようになってきているのだ。思いと違うなと思えるところも、当たらずとも遠からず成ってきている。アイデアとは物を作ることで洗練させていくもの・・・実感です。

確かに当初書いた仕様書とは違うことは多々あっても、その文章を書いた心は変わらない。実際そこへ向けてプログラムは近づいていっているようだ。歩みはのろいけれど、着実に人工知能の本質に迫ってきている。それを実感するこの頃でもある。頑張りたい。

おわり