

随筆：「自律成長型人工知能」

4月13日から科学技術週間が始まりました。そして、4月18日は発明の日。私も、なんとか頑張って新しい目標に向かって始動し始めました。言語を自律で習得していく人工知能を開発していくことにしたのです。これは、自然言語処理、画像認識、思考システム、すなわち未夢・香澄プロジェクト、青葉・若葉プロジェクトを包含し、価値を押し進めるものです。今までの自然言語処理ですと、単語や文法を人手で辞書に持たせていかねばなりませんでした。また、青葉・若葉プロジェクトを実現するとしても、最終目標は自然言語処理と画像認識処理を統合することでしたので、ここは一気に人間並みの機能を実現することを目指そうと考えたわけです。人生は一度きり。後悔はしたくありません。精一杯の事を成していこうと思ったわけです。音声認識もやらねばなりませんかね。人間とやりとりしながら言語を自律で習得していくシステムでなくてはなりませんから。

こんなことを考え始めたのは、認知言語学の本を幾つか読んだことによります。今まで、考えてきたことのコアが認知言語学に記述されているようなことなのです。認知言語学をもっと数学的というか、オントロジーを基本に考えていくと、そのままコンピュータで実現できるような科学技術になるのではないかと感じたのです。認知科学を極めつつ、もの作りしていこう、それがこれからの私の生きる道ではないか、そう思うようになってきました。認知科学をコンピュータ技術の言葉で徹底的に表現していく。そうする時期に来ているのだと思うのです。

さて6年前、会社を離れるとき同僚の方々から饞別を貰ったのでした。図書券を貰ったのです。今回、これまでの研究をまとめた「4月7日の贈り物」を公表してから、認知言語学を勉強しようと本を買いまくりました。その資金にいただきました。つまり、会社の皆様のご厚意で新しい目標が設定されたことになります。感謝です。

これからは私生活がちょっと苦しくなりますが、今まで以上に頑張りたく思います。未夢・香澄プロジェクト、青葉・若葉プロジェクトの成果がありますし、認知科学の大きな成果を勉強できますし、これからの楽しみです。0からではありません。見通しは良いのです。もう、闇雲にもの作りしなくてはならないという苦しいものではありません。技術は大分整理されてきていると感じます。もう一押しです。もう一押しで輝かしい未来が開かれるのです。

ということで、ご挨拶で、いまどんなことを考えているか語ってみたいと思います。自律成長型人工知能はどうあるべきかということ。基本はオントロジーとオンデマンド解析と、全件内容検索技術、パターン発見技術という超並列コンピューティング技術になるだろうと。そのところを少し説明していきたいとおもいます。

カテゴリーシステム

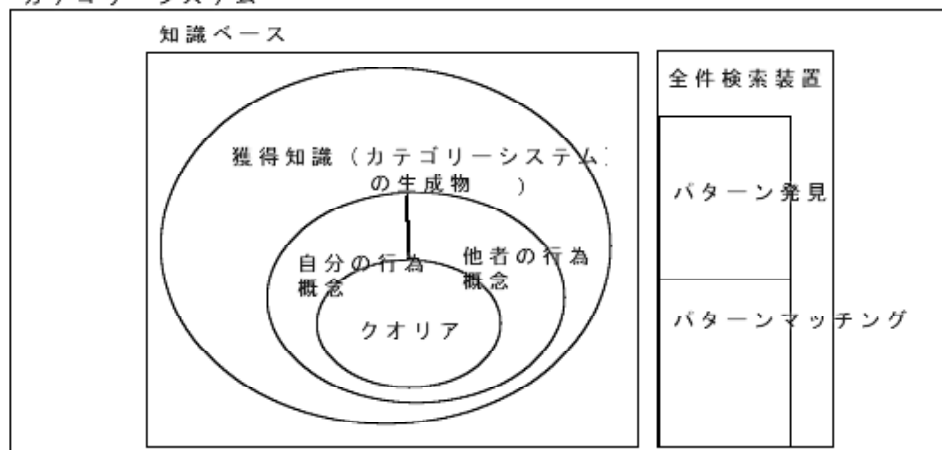


図 1. 自律成長システムの基盤機構

自律成長システムは先ず成長の基盤となるデータシステムが必要です。結晶の種になるものですね。それは身体です。身体は外界を認識する基盤ですから、感覚情報を評価する基本的なハードウェアを備えています（クオリア）。また、行為を認識するために基盤となる概念も人工知能は生得的に持っていません。そして、他者の行為をこの自分の行為に翻訳して解釈できないと物まねによる学習ができません。教授者の意図を学習者が認知出来なくてはなりません。他者の行為概念ですね。これが重要です。それは、このミラーニューロンシステムが必要ということです。それはまた一つのオントロジーシステムでしょう。

そうして、パターン認識をしながら知識を獲得していきます。パターン認識はパターン発見装置とパターンマッチング装置群によって実現されます。いずれも、データを全件検索する超並列処理装置が基盤技術として必要になります。

人工知能は超並列処理で実現していくしかありません。身体の入力を実時間でこなすのですから当然です。そして、身体の入力はカテゴリーシステムに送られて、そのデータを興奮させます。発火ですね。なにか特徴データを選んで処理していくという、従来のコンピュータシステムでは自律成長型人工知能は実現できません。なにが大事かということは自律的に学習して身につけていく物ですから、全ての入力データを扱えるアーキテクチャが必須なのです。

図と地などの基本的な価値評価をする評価システムも必要です。地としては枠組みがなるもので、図は注意が起きる場所・・・行為者のことになります。そんな規則も行為の概念知識として身体性が備えるデータです。

コアとなる身体性データの上にカテゴリーシステムが構築されるのですが、カテゴリーシステムは認知言語学が語るように、発火される身体性データ群の特定のパターンの頻度を係数して、高頻度のものをカテゴリーにしていくということが基本になります。そうして、カテゴリー群の中で、さらに共通の項目（パターン）を持つ物がスキーマとして抽象カテゴリーになって、上位に設定されていくことになります。知識ベースはそのようにして、ネットワークとして充実されていく物なのです。

発火もパターン発見技術も全件検索技術が基本となります。また、発火は文脈を形成しますから、文脈により様々なプロセスを平行起動していくことにも全件検索技術が重要とすることになります。発火はいかに迅速に入力データにマッチングする知識データを全てオンにできるかという処理です。知識データを全件検索しながら、入力データとマッチングしていくこととなります。パターン発見も既存の全パターンと発火したデータの一致度をマッチングしていく処理です。文脈も発火した知識データのことですから、その全てと所定の条件とが一致するかどうかをマッチングしていく処理が基盤となります。

結局、人工知能は超並列処理として実現していくしかないのです。なんでもかでも並行処理に帰着されますから。ただ、アルゴリズムは単純なものが多いようです。単純なプロセスが幾つかあって、それが超並列の無数のスレッドとして実行されるのです。

このような超並列機構をRTミドルウェアに組み込んでいくこととなりますが、専用ハードウェアで超並列を実現するか、メインCPUをメニーコアにして実現するかはシステムの環境によります。どんなシステム環境にも対応できるように人工知能は作られているべきでしょう。RTミドルウェアの各コンポーネントが環境を判断して処理をスイッチするようにすればいいと思います。そんなイメージを下図に示します。

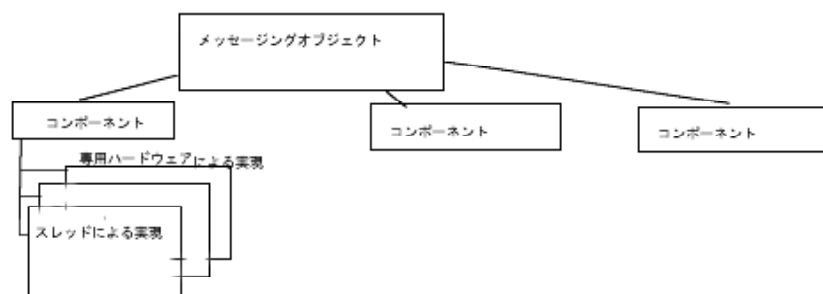


図 2. 超並列機構の実現

オンデマンド処理の一つに、バックトラッキング法というものがあります。通常はデフォルト処理をしていきますが、処理結果に欠陥を発見すると、再度別の解析結果を探してそれをもっと処理を試みる。これを文解析に適用することを考えてみますと次のようになります。文解析は同時に次の処理を文頭からやっていくのでした。

- ・分かち書き化
- ・係り受け解析
- ・格の所属解析
- ・格の意味解析

(例文)にはにはにわにわとりがいる。

この例文は結構難しいですが、分かち書き化として単純に最長一致法を適用します。デフォルトとしてはそんな処理を選択するのが素直でしょう。そして、係り受けも格の所属

も直後の名詞とか動詞を選択していく。しかし、係り受けや格の意味解析で齟齬が生じたときは、再度、分かち書きを変えるか、格の所属を変えていきます。そうして、バックトラックします。バックトラックの管理は、それまでの失敗した解析結果をリストにしておくことで、同じ処理を何遍もすることを避けることになります。

おわり