

人工知能はパターンマッチングの処理がすごく多いです。高速に実現しないと、リアルタイム処理に深い影を落とします。図面の線画解析結果から、意味ネットワークに落とすところとか、日本語の分かち書き化から助詞と動詞を中心としたパターンマッチングから、格を考慮した意味ネットワークに落とすところは、まさにパターンマッチング技術の工夫点がいっぱいあります。今回は、ここを考察していつてみたいと思います。

パターンの知識要素はストレージのRDBであっても、インメモリのオブジェクトネットワークに成っていた方が高速化には有利です。いずれ、オブジェクトネットワークでパターンマッチングしていかねばなりません。データには構造があるからです。といっても、そんなに深い構造でなく、2階層くらいの木構造ですが。

(例文)

- (1) 私は図書館で本を読む。
- (2) 私は眼鏡で本を読む。
- (3) 私は必至で本を読む。
- (4) 私はそれを本で読んだ。

.....

同じ、「～で～読んだ」パターンの文群です。精確に格を判定するにはどうしたらよいでしょうか。

解釈には知識パターンが必要です。基本的に、動詞と助詞群と助詞にどんな意味を持った名詞が支配されているかをもって、格を判定していきます。助詞の組み合わせで、それぞれの助詞の格が決定されていく(あいまいのままのこることもあります)わけです。しかも、助詞がどんな意味の名詞を支配しているかかかわる問題ですので、すごく複雑です。もう、知識をもって対応してくしかないわけです。アルゴリズムで対応するのではなく、過去の経験から助詞の格を推論していくことになるわけです。

ということで、(例文)のようなものは、すべて知識に持っていて、できればそれを抽象化し、たとえば、「図書館」は「location」というような意味になっていれば、

- (5) 私は駅で本を読む。

などの文の格は「駅」が「location」なる意味を持っていたら、そのままで推測できるわけです。

そうして、更なる問題は、知識は無数にあって、それと入力文とのマッチングは重いなということです。それを少しでも速く行うには、構造を含めたマッチングでなくて、先ずは意味ネットワークにどんな名詞があるかを調べていくということがあると思います。で

できれば、名詞と助詞の組み合わせ情報をどの意味ネットワークが担っているかをあらかじめ知ることは有利です。それには、名詞リストを意味ネットワークの情報として持っていて、まずは、このリストをマッチングさせていく。そんなことですね。

マッチングは完全なマッチングということは稀でしょう。入力文は欠落語を多分に含んでいます。だから、マッチングは全ての知識について「確信度」でマッチング状況が表示されることになるはずですが。最後に、もっともマッチングしているものを選ぶ・・・上位何位かを選んで、文脈処理で最終結果を得ていく、などの処理になっていくと思えます。

このようにしますと、欠落語にも対応できます。意味ネットワークのオブジェクトとパターンマッチングの名詞オブジェクトとが対応づけられるから、欠落語はあとの文脈処理で埋め込めばいいのです。文になにが足りないかも知識パターンから得られる貴重な意味解析の手がかりです。

あとは、機械学習機構ですね。ある文が与えられて、完全にマッチングできない状況ですと、もっとも近いパターンで推論して、表示します。良好と人が判断したら、そのパターンを記憶します。人に間違いを指摘されたら、直して、記憶していく。抽象化もデータマイニングとかテキストマイニング手法で行っていく。そうすれば、結構楽しいロボットができるかも知れません。結構、賢いと評価できるロボットになるのではないのでしょうか。

おわり