

イラストの認識を考えて行ってみたいと思います。パターン認識を特徴点などのオブジェクトとそのオブジェクト間の関係情報を持って投票法で解法していくというのが技術でした。更に、その上で、人物が本を抱えて立っているとか、誰かに本を渡しているとか、そんなシーンの認識が展開されるでしょう。その時大事になるのが、自然言語認識での技術です。自然言語認識では、文を、格とオブジェクトをパラメータにもつ、動詞をコマンドとするリストに落とし込むということでした。更に、その認識では、オントロジー記号によるコマンド群に落とし込み、たとえば、「人 a が人 b に本を渡す」では、

- (1) (belong,agent_case,book,to_case,human_a)
(have,agent_case,human_a,direct_object,book)
- (2) (move,agent_case,book,from_case,human_a,to_case,human_b)
(reave,agent_case,book,from_case,human_a)
- (3) (belong,agent_case,book,to_case,human_b)
(have,agent_case,human_b,direct_object,book)

というような、時間を追ったコマンド表現として記述されるでしょう。これが、シーンの意味理解の為の文脈知識となります。

この文脈知識を画像認識結果と対応付けるデータがあります。画像認識の結果から、人間の認識では、手や足、腕の状況といったことも解析する必要があります。人物を認識しながら、手足腕の認識とそれらの状況の認識をします。それは、特徴点というような画像のオブジェクトでなくて、人間の持っている概念のオブジェクトであるということです。それは、オブジェクトの認識で、投票法で実現して行きます。そのことは、人物というものが認識されれば、その構成部品が何であるかという絞込みは「常識」データが解決するでしょう。あとは、そのデータ内部で、どれにいちばん近くて、その形状は、属性はなにかと解析していくことになるのです。

手が本を持っているような状況では、腕が水平で、その上に長方形が描かれたイラストに成っている場合が多いでしょう。腕と長方形の接続の連続性から、「腕は本を抱えている」という事象に投票していく。シーンとしては、その連想から、「本を持っている」というコマンドに投票します。その状況は、本と人間に関係する状況という枠組みで、シーンの連想が起きて行くでしょう。それで、シーンの絞込みをしていきます。

そうして、最後に、文脈知識の投票して行って、イラストの認識は完成していきます。

このように、すべては、データの絞込みと絞り込んだ中での投票による、連想結果の選別という作業になっていくことが分かりました。

おわり