

自然言語処理入門

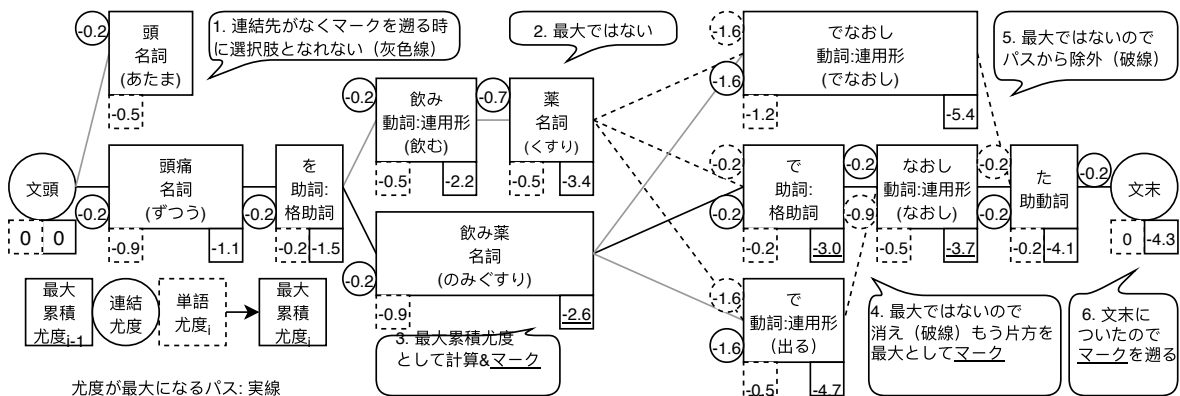
岸山 健 (31-187002)

Oct. 29, 2018

1 課題

以下に示す単語辞書と連結可否行列を用いて、「頭痛を飲み薬でなおして」を形態素解析せよ。

まずは単語辞書を用いてラティスを作成する。更に各ノードに対して2つの箱を用意し、一つにはその単語の尤度、もう一つには最大累積尤度を入れる。また、エッジには接続尤度を置く。尤度の積が最大のものを求めたいが、確率の積は小さくなるため対数変換した和を求める。そのために、与えられた単語辞書と連結可否行列にある確率の対数を取り尤度とし、それぞれの箱とエッジに割り振る。その結果が以下のラティスとなる。



このラティスにはビタビ・アルゴリズムを適用しており、ノードの右下にある実線で書かれた箱には最大累積尤度が格納してある。最大累積尤度を持つノードは真偽値の真を格納する。(つまり全てのノードは `isPrevious` のような、bool 値を持つ変数が必要である。) このラティスでは、各ノードの右下にある最大累積尤度の下線部を引いて表した。ビタビ・アルゴリズムを適用し、現在処理しているノードが文末であるならば、現在のノードからリストに格納していく。前のノードの候補が2つ以上ある場合は `isPrevious` を確認し、真であるノードをリストに格納する。この作業を文頭まで続け、最後にリストを逆順にすればパスが求まる。このラティスでは尤度が最大となるパスを黒の実線で示した。