言語モデルとは単語や文章の出現確率を数値化したものです。すなわち、単語や文章全体がある確率分布に従っていると仮定します。そして、大量のテキストデータから学習をおこない、ある単語の後に続く単語がどのぐらいの確率で出現するかの予測などを行います。

自然言語処理の分野ではコンピューターに人間の言葉を理解させるために言語モデルが使用されます。自然言語処理のタスクは主に、形態素解析、構文解析、意味解析、文脈解析です。形態素解析では、文章を最小の単位である単語に分割します。例えば、「満天の星と月が美しい」を形態素解析すると「満天/の/星/と/月/が/美しい」となります。構文解析は、単語同士の関係性を解析する作業です。例えば「満点の星と月が」の部分だと、「星と月が」満天なのか、「満天の星」と「月」なのかで、異なる捉え方ができます。構文解析では、そのような考えられる捉え方をすべて列挙します。意味解析では、単語間の関係を調べる作業です。上の例だと、星は複数あり、月は1つのみだから「満天の星」と「月」が正しいと特定できます。

次に、言語モデルで出てくる手法について説明します。ここでは、n-gramモデル、隠れマルコフモデル、ニューラルネットワークモデルについて説明します。

n-gramモデルは、マルコフモデルに基づいて単語列をモデル化します。ここで、マルコフモデルとは、将来の状態が現在の状態のみに依存する性質（マルコフ性）を持つモデルのことをいいます。例えば、単純な例として天気がマルコフ性を持つことを考えます。このとき、明日の天気（晴れ、くもり、雨）は、今日の天気のみに依存することになります。

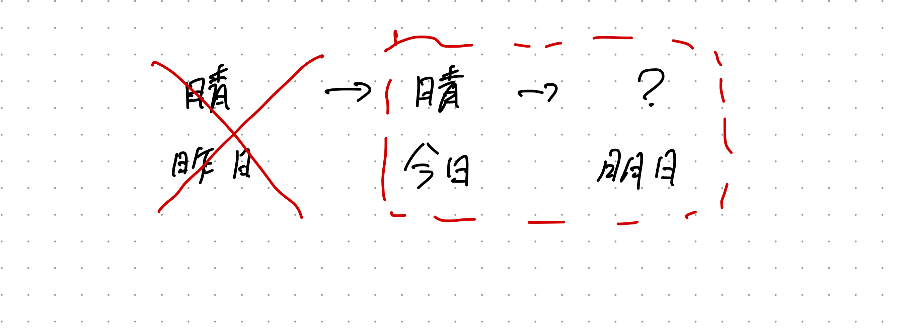


図 1 マルコフモデルの例

n-gramモデルでは単語列内の次の単語の確率が、先行する単語の固定サイズのウィンドウにのみ依存すると仮定することで単純化します。すなわち、前の単語を1つだけ考慮する場合はbigramモデル、2つの単語を考慮する場合はtrigramモデル、先行するn-1単語を考慮する場合はn-gramモデルと呼ばれます。例えば、bigram言語モデルは、「I know the old man.」という確率を次のようにモデル化します。

ここで、<s>, </s>は文章のはじめと終わりを意味する特別な文字列です。これらの条件付き確率は、テキストコーパストを用いて推定することができます。

隠れマルコフモデルは、マルコフモデルのうち状態が観測されない状況を加味したモデルのことをいいます。前の天気の例では、状態（晴れ or 曇り or 雨）が観測できるため、状態=観測結果となります。なので、これは隠れマルコフモデルとは言えません。隠れマルコフモデルの例としては以下のようなものが挙げられます。状態が「元気」、「風邪」であり観測できず、観測できる出力結果を「体温」とします。そして、状態と出力に何かしらの確率（出力確率）を仮定します。

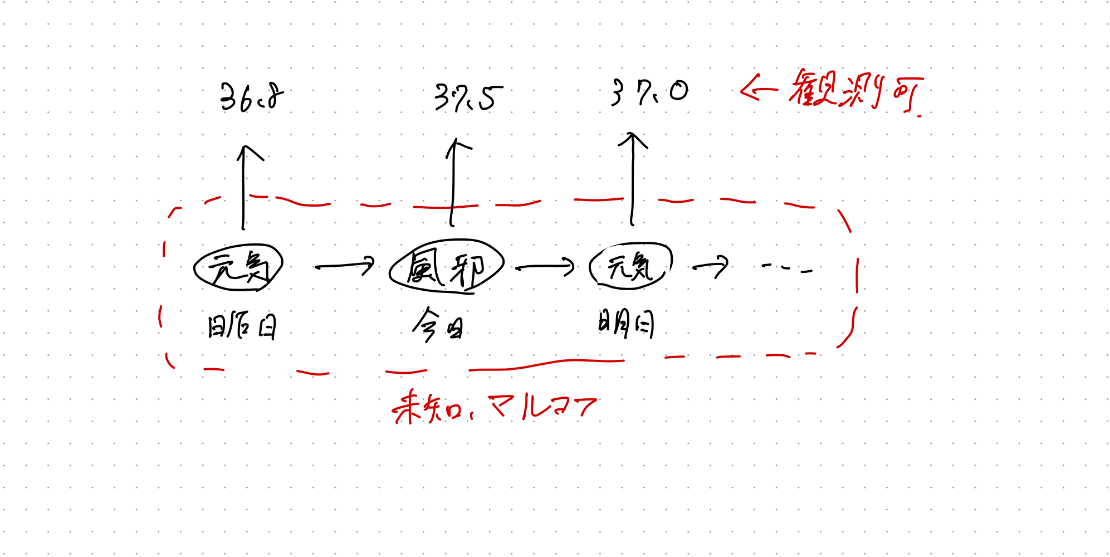


図 2 隠れマルコフモデルの例

隠れマルコフモデルのパラメータは、遷移確率（状態が変わるときの確率分布）と出力確率です。隠れマルコフモデルでは、この2つのパラメータを推定することができます。

この隠れた状態を推定できるという性質が、自然言語処理のタスクに活用できます。例えば、形態素解析では、状態を品詞とすることで、あたえられた文章の中の単語がどの品詞になるかを効率的に推定することができます。また、音声認識では、状態を単語列とし、出力を音声データとすることで、与えられた音声から単語を推定することができます。このようにして、隠れマルコフモデルは自然言語処理のタスクで活用されます。

ニューラルネットワークモデルは、単語や文章の解析にニューラルネットワークを用いるモデルです。訓練データとしてテキストが用いられるのですが、訓練データで用いられる語彙の数が多いと、単語間の組み合わせの数が指数関数的に増加します。そのため、ニューラルネットワークモデルでは、単語語句を実数ベクトル空間上に移す（埋め込む）ことを行ないます。

単純な埋め込みとして、one-hot表現があります。例えば単語全体の集合が{りんご、みかん、なし、ぶどう}だとして、りんごをベクトル表現することを考えます。このとき、4次元のベクトルの各成分に各要素（りんごならば、第1成分を1とするなど）を割り当てることで、単語のベクトル表現ができます。

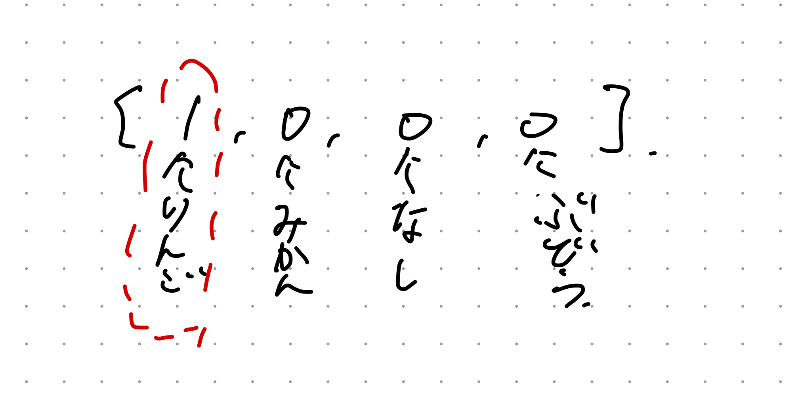


図 3 単語のone-hot表現

しかし、この表現だと、各単語間で類似度がはかれないという問題点があります。それに対する埋め込み方法として、分散表現があります。

分散表現は、各単語を50~300次元の実ベクトルで表現します。このように表現することで、各単語の類似度を簡単に図ることができます（たとえば内積を計算）。

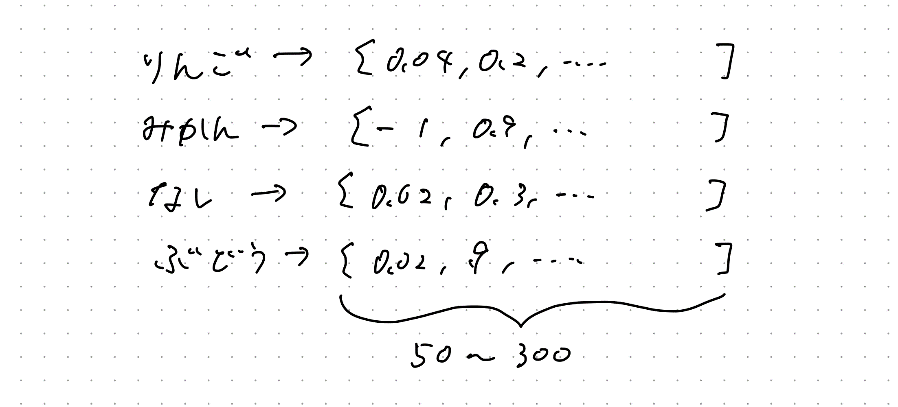


図 4 単語の分散表現

分散表現の有名な手法として、Word2vecが挙げられます。Word2Vecの場合、意味的に似ている単語はベクトル空間上でも近くに位置するように設計されています。これにより、「王 - 男 + 女 = 女王」といった単語間の意味的な関係を表現することが可能になります。この分散表現をニューラルネットワークモデルに利用することで、より精度の高い結果、複雑な表現を得ることができます。

（参考）

[自然言語処理（NLP）とは？できることや活用事例をわかりやすく解説｜ITトレンド (it-trend.jp)](https://it-trend.jp/textmining/article/124-0027)

[【技術解説】マルコフモデルと隠れマルコフモデル - ミエルカAI は、自然言語処理技術を中心とした、RPA開発・サイト改善・流入改善レコメンドエンジンを開発 (mieruca-ai.com)](https://mieruca-ai.com/ai/markov_model_hmm/)

[言語モデル - Wikipedia](https://ja.wikipedia.org/wiki/%E8%A8%80%E8%AA%9E%E3%83%A2%E3%83%87%E3%83%AB)

[単語の埋め込み - Wikipedia](https://ja.wikipedia.org/wiki/%E5%8D%98%E8%AA%9E%E3%81%AE%E5%9F%8B%E3%82%81%E8%BE%BC%E3%81%BF#cite_note-10)

[なぜ自然言語処理にとって単語の分散表現は重要なのか？ #Python - Qiita](https://qiita.com/Hironsan/items/a58636f946dd51f670b0)