

エルマンネットを拡張した語系列予測モデルによる文法特徴の獲得

篠沢佳久[†]

慶應義塾大学 理工学部 管理工学科[†]

1. はじめに

エルマンは、単純再帰結合型ネットワーク（エルマンネット）を用いて、言語の獲得過程のモデルを提案した[1][2]。これは、エルマンネットを用いて現時点のみの単語を提示し、次単語を予測する学習（次単語予測学習）を繰り返すことによって文法特徴の獲得の可能性を示した。エルマンネットによる文法獲得の学習は、文法規則に制約を設けた人工文を対象とし、特に含まれる単語についてはその個数も限定されている。そこで本研究においてはエルマンネットを拡張し、単語数の増加に対応が容易な語系列予測モデルを提案し、次単語予測学習を通しての評価を行なう。

2. 提案モデルの構造

エルマンネットは、多層パーセプトロン型を基本構造とし、入力層、外部からの入力値を処理する入力層と直前の中間層の活性状態がフィードバックされる文脈層の二つから構成される。次単語予測学習をエルマンネットによって行なう場合、入力層および出力層に総単語数分のニューロンを用意し、任意の一単語と特定の一ニューロンの対応づけを行なう。そして文章を単語ごとに区切り、現時点の単語と対応する入力層のニューロンには1を、それ以外は0というような単語表現で順次入力する。文脈層については、中間層と同じニューロン数を用意しておき、直前の中間層の値（内部表現）をそのままフィードバックして入力する。出力層は、次単語と対応するニューロンのみが発火し、それ以外は発火しないように結合係数の修正（学習）を行なう。従って構造上、単語数が増加するに従い学習が困難となり、新たな単語を追加学習することは容易ではないものと予想される。そこでエルマンネットを次の三点から拡張することによって単語数の増加に対応が容易な語系列予測モデルの実現を試みた（図1）[3]。

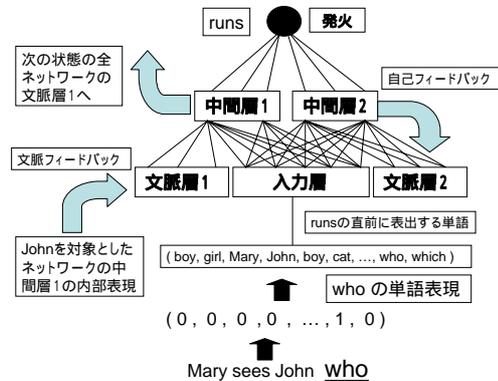


図1 提案モデルの構造

- 出力層はニューロンを1個と制限し、特定の単語と対応づける。すなわち特定のネットワークは対応づけされた一単語のみの予測を行なう。次単語予測学習でN個の単語を対象とする場合、N個のネットワークが必要となる。
- ネットワークの入出力を学習用の例文中に表出する単語の接続情報（パイグラム）を利用し制限する。すなわち学習例文を用いて単語iの直前に表出する単語集合 S_i を求める。そして単語iに対応したネットワークの入力層は単語集合 S_i のみを入力対象とする。
- 文脈層へのフィードバックについては、直前の中間層の値を戻す自己フィードバックと正解となる単語を対象としたネットワークの中間層の値を戻す文脈フィードバックの二種類とした。さらに中間層と文脈層を二つに分け、蓄積される単語の履歴情報と文脈情報が混在して学習しないために自己フィードバックと文脈フィードバックによって再入力する中間層と文脈層を対応づけた局所結合型構造とした。入力層は全ての中間層と結合している。

3. 提案モデルによる次単語予測学習

単語数をN個含む文章を学習対象としたN個のネットワークの状態を想定する（初期状態は $N=0$ ）。提案モデルを用いた次単語予測学習のアルゴリズムは下記の通りである。

学習用の例文数をJ個とする。各入力文章 $L_j = \{w_j(0), w_j(1), \dots, w_j(t), \dots, w_j(T)\}$ （例文数 $j = 1, 2, \dots, J$ ）を単語単位 $w_j(t)$ に区切る。

Grammar acquisition by a model of word sequence prediction by improved simple recurrent networks

[†]Yoshihisa Shinozawa, Faculty of Science and Technology, Keio University

初期状態を $t=0$ とする。

全ネットワークの中間層を初期化する。

次単語 $w_j(t+1)$ を対象としたネットワークが存在しない場合、単語 $w_j(t+1)$ を対象とするネットワークを生成する ($N=N+1$)。そのネットワークの入力層は単語 $w_j(t)$ を対象としたニューロンのみとする。また次単語 $w_j(t+1)$ を対象としたネットワークにおいて入力層に単語 $w_j(t)$ を対象としたニューロンが存在しない場合は、単語 $w_j(t)$ を対象としたニューロンを入力層に追加する。

各ネットワーク k ごとに単語 $w_j(t)$ を単語表現 $V_{jk}(t)$ に変換し入力層に入力する。

直前の単語 $w_j(t-1)$ に対応したネットワーク m の中間層 1 の値 $H_m(t-1)$ を全てのネットワークの文脈層 1 にコピーする (文脈フィードバック)。また各ネットワーク k ごとに直前の中間層 2 の値 $P_k(t-1)$ を文脈層 2 にコピーする (自己フィードバック)。

出力層は次単語 $w_j(t+1)$ に対応したネットワークは発火し、それ以外は発火を抑制するように誤差逆伝播則によって学習を行なう。

単語 $w_j(t+1)$ が文末でなければ、 $t=t+1$ としへ戻る。文末であれば $j=j+1$ としへ戻り、別の例文で学習する。

以上、例文の個数分だけ から を繰り返す。

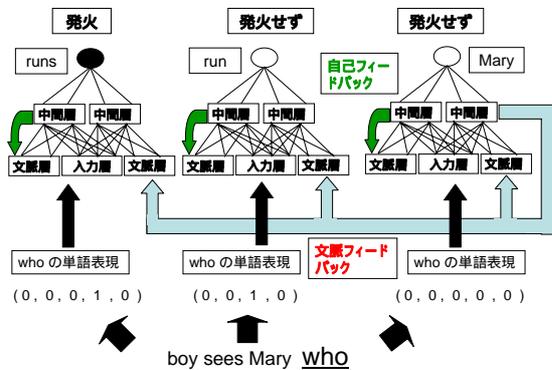


図 2 提案モデルによる次単語予測学習

4. 評価実験

エルマンのシミュレーションで使用された人工文の生成規則[2]を拡張し 440,000 文の例文 (単語数 89 個, 単語のカテゴリ 25 種類, what などを文頭にした疑問文も含む) を作成し, 半数の例文を用いて次単語予測学習を行なった。提案モデルにおいてネットワークは単語数分必要なので 89 個, 各ネットワークは 4 層型とし, 圧縮層は 20 個, 二種類の中間層と文脈層はそれぞれ 64 個 (合計で 128 個) とした。そして通常

のエルマンネット (4 層型, 入力層と出力層は 89 個, 圧縮層 20 個, 中間層と文脈層 128 個) を用いて同様な条件で学習を行なった結果と比較する。評価指標にはエルマンの利用した COSINE[2], 正解となる単語の平均予測順位[3]を用いた。図 4 に学習回数ごとの評価指標の推移状況を示す。また学習停止後, 未学習の例文に対し次単語予測を行なった結果を表 1 に示す。

図 3 学習回数ごとの評価指標の推移

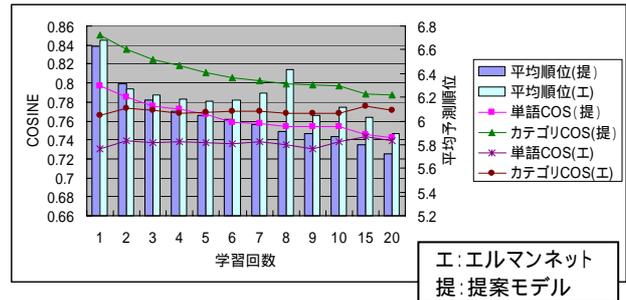


表 1 未学習の例文に対する予測結果

	予測順位	単語 COS	カテゴリ COS
エルマン	5.90	0.738	0.771
提案モデル	5.95	0.741	0.787

学習停止後, 単語カテゴリーに対する COSINE の指標が提案モデルの方が高い (表 1)。これについて分析したところ, 疑問文中における文末「？」などの予測において, 提案モデルは長い文章の場合でも正確であり, 文脈フィードバックによって過去の文脈情報をより正確に蓄積していることが分かった。

5. まとめ

本研究においては, エルマンネットを拡張した語系列予測モデルを提案し, 次単語予測学習による評価を行なった。今後は提案モデルにおいて, 中間層の活性状態を分析し, 文法特徴の獲得についての解析をすすめていく予定である。

参考文献

- [1] Elman, J. L., "Finding Structure in Time", Cognitive Science, 14, pp.179-211 (1990)
- [2] Elman, J. L., "Distributed representations, simple recurrent networks, and grammatical structure", Machine Learning, 7, pp.195-224 (1991)
- [3] 篠沢佳久, 単語の接続情報を利用した単純再帰結合型ネットワークの拡張, 情報処理学会研究会 NL175-12, pp.79-84 (2006)