

P4-6 品詞タグ付け ニューラルネットワークの深層化

坪井 祐太 <yutat@jp.ibm.com>

日本アイ・ビー・エム株式会社 東京基礎研究所

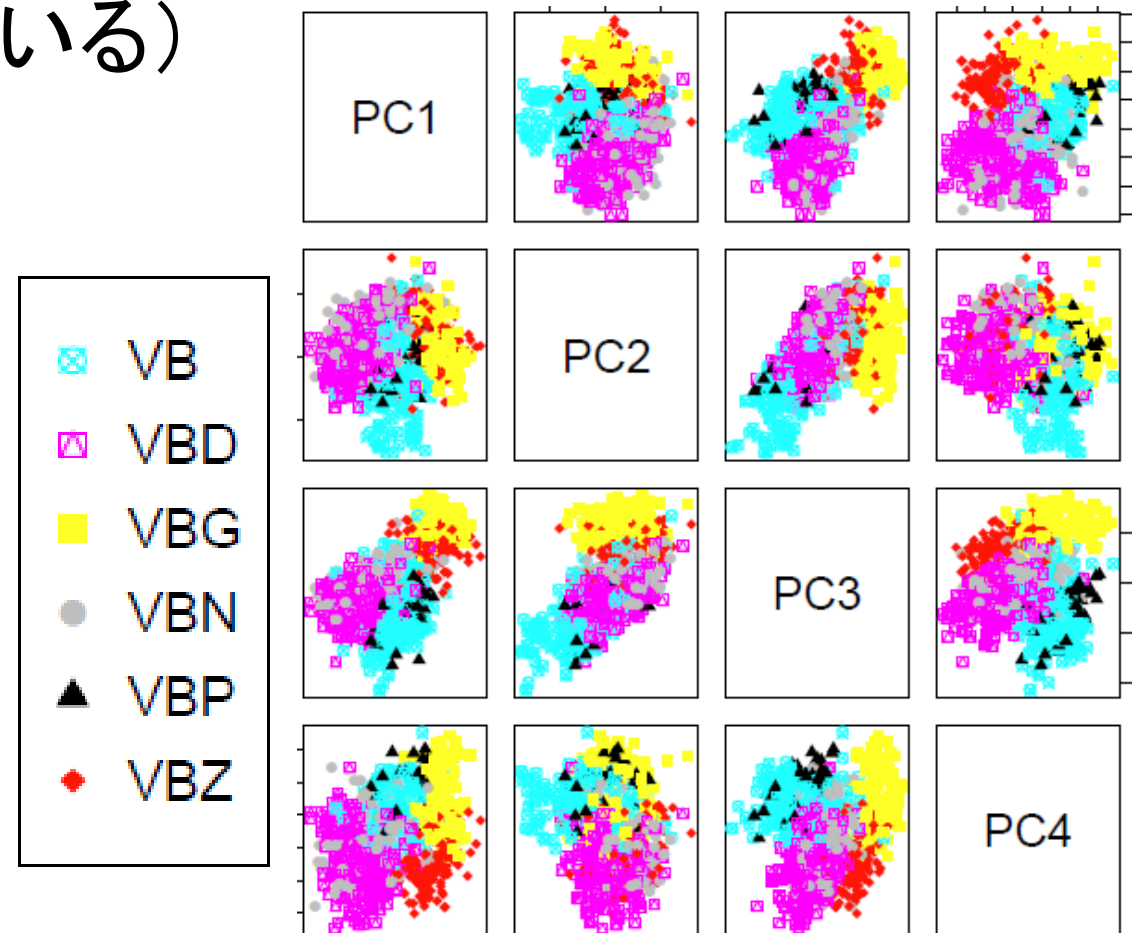
概要

- ニューラルネットワークに基づく英語品詞タグ器 [Tsuboi, 2014] を深層化
 - 最高精度達成: Penn Treebank (WSJ)
97.51% (ours) vs. 97.50% [Søggard, 2011]
- 同程度の精度のまま高速化と省メモリ化
 - パラメータ数: 約30%に減少
 - 解析速度: 約4倍に高速化

ニューラルネットワークに基づく 英語品詞タグ器 [Tsuboi, 2014]

- 連続値特徴量に対してニューラルネットワークで非線形変換を学習
 - 2値特徴量に比べて直感の働きにくい連続値の組み合わせ特徴量に暗に対応
- 既存研究の2値特徴量(線形)と連続値特徴量(非線形)を組み合わせ高精度を達成

開発セットの動詞に対応するニューラルネットワークの隠れ変数ベクトルを主成分分析した結果(品詞クラスタを学習できている)

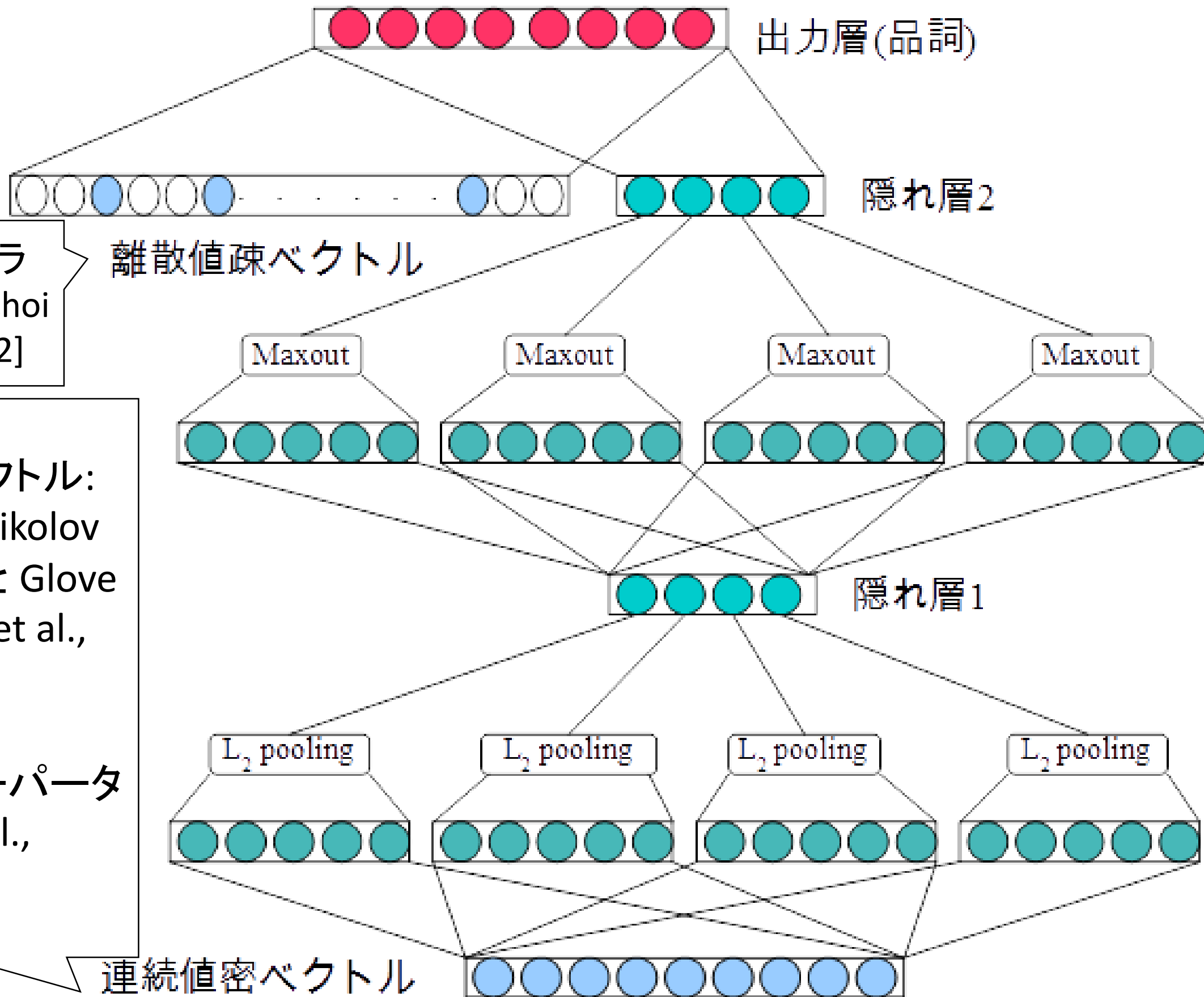


| | テストセット精度 | |
|-----------|----------|--------|
| | 全単語 | 未知語 |
| 2値のみ(線形) | 97.12% | 87.28% |
| 連続値のみ(NN) | 97.40% | 91.23% |
| 両方(線形&NN) | 97.51% | 91.64% |

ニューラルネットワークに基づく 英語品詞タグ器 [Tsuboi, 2014] 詳細

- 予測履歴も特徴量とする決定的タグ付け器
 - 2値特徴量: 単語Nグラム、予測品詞/頻出品詞Nグラム、接頭辞、接尾辞、文字種など [Choi and Palmer, 2012]
- 学習しながら訓練事例を生成 [Goldberg and Nivre, 2012]
 - 学習中のタガールの予測履歴を訓練事例の特徴量として追加
 - 誤った予測履歴も考慮して学習するため誤差伝播を防げる
- FTRLProximalオンライン学習 [McMahan, 2011] とAdagrad [Duchi et al., 2010]を併用
 - 多クラスヒンジ損失 & L1/L2 正則化 (L1正則化は線形モデルのみ)
- ランダムなハイパーパラメータの組み合わせを開発セットで評価しモデル選択 [Bergstra and Bengio, 2012]
 - 対象: パラメータ初期値・初期値範囲・モーメントムON/OFF・学習率・正則化パラメータ・正則化を開始するエポック数

提案法: 線形モデルと深層ニューラルネット



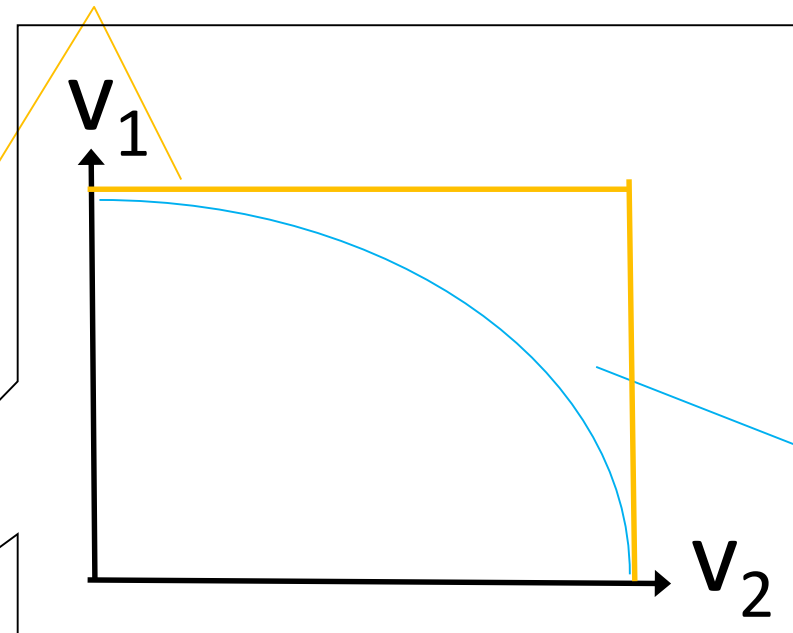
単語や品詞Nグラム・文字種など [Choi and Palmer, 2012]

- 5342 次元
- 埋め込みベクトル: word2vec [Mikolov et al., 2013]と Glove [Pennington et al., 2014]
 - 品詞分布
 - 係り受けスーパータグ [Ouchi et al., 2014] 分布
 - 隣接語分布

Maxout Networks [Goodfellow et al., 2013] & 正規化 L_p プーリング [Gulcehre et al., 2014]

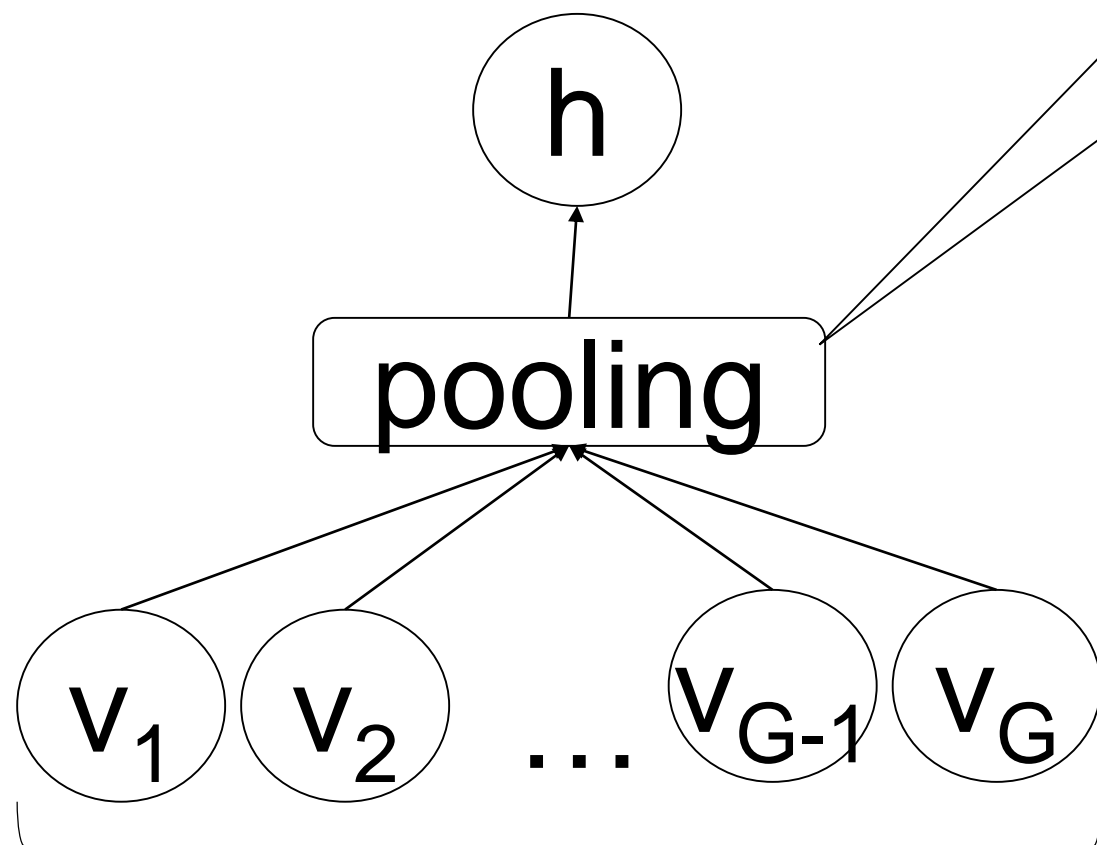
Maxout Networks

$$h = \max(v_1, v_2, \dots, v_G)$$



Normalized L_2 Pooling

$$h = \left(\frac{1}{G} \sum_{j=1}^G |v_j|^p \right)^{1/p}$$



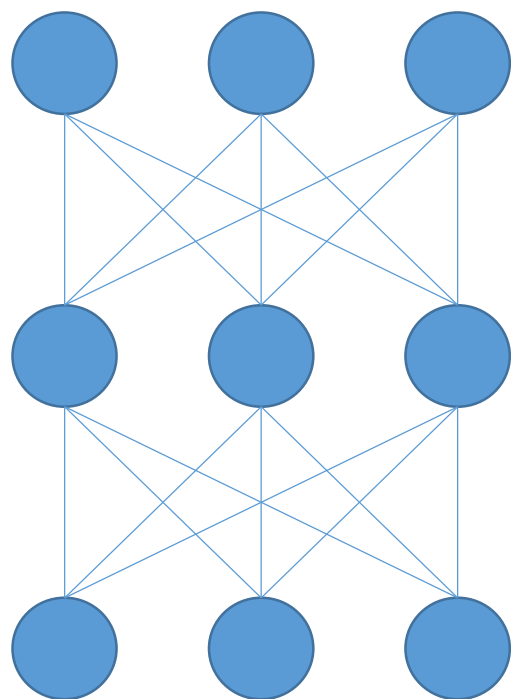
線形変換
 $\mathbf{v} = \boldsymbol{\theta}^T \mathbf{X}$

要素数Gのグループ

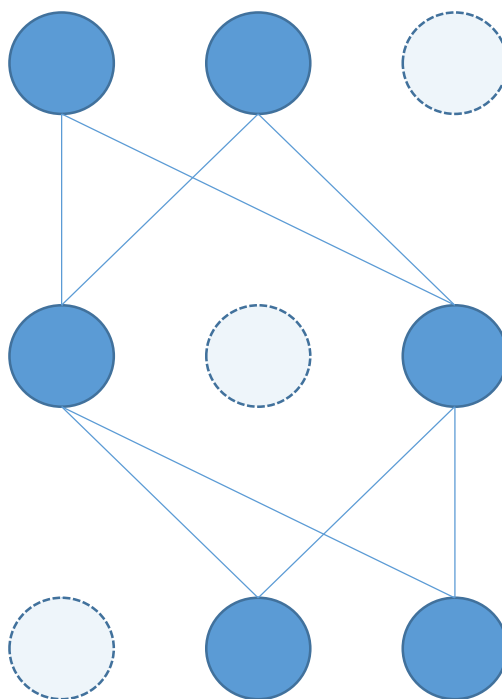
Dropout [Srivastava et al., 2014]

- 訓練時にドロップアウト確率 p で隠れ変数 h を0に置換
 - 事例(ミニバッチ)毎に異なるネットワーク構造を評価・更新していることに相当過学習防止効果
 - テスト時には学習結果パラメータを $(1-p)$ 倍することで、擬似的に複数のネットワークの幾何平均で予測していることに相当
- 過学習防止効果

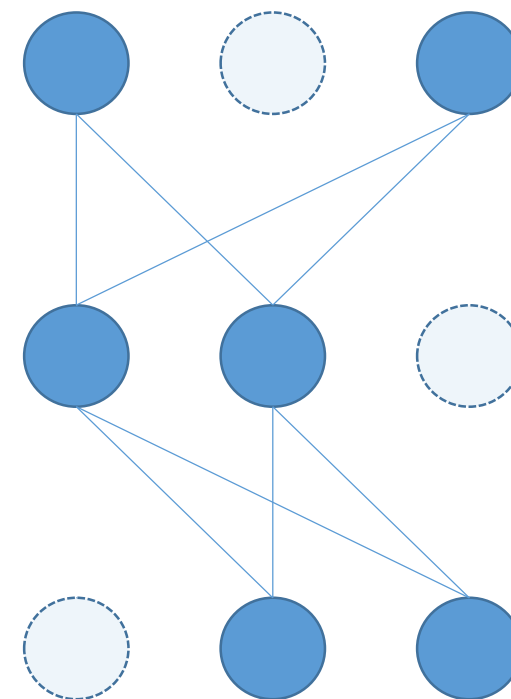
ドロップアウトなし



ドロップアウト例1



ドロップアウト例1



実験設定

• データセット

- Pennツリーバンク
標準的分割
- 45品詞タグ

| データセット | 文数 | トークン数 | 未知語数 |
|-------------|--------|---------|-------|
| 訓練 (0-18) | 38,219 | 912,344 | 0 |
| 開発 (19-21) | 5,527 | 131,768 | 4,467 |
| テスト (22-24) | 5,462 | 129,654 | 3,649 |

• ハイパーパラメータは開発セット評価結果を基に選択

| ハイパーパラメータ | 層 | 候補 |
|-------------|---|----------------------|
| 隠れ変数の数 | 1 | {4, 8, 16} |
| | 2 | {8, 16, 32, 64, 128} |
| 活性化関数のグループ数 | 1 | {4, 8, 16} |
| | 2 | {4, 8, 16, 32} |
| ドロップアウト確率 | 0 | {0.0, 0.2, 0.4} |
| | 1 | {0.0, 0.5} |
| | 2 | {0.0, 0.5} |

実験結果

- 入力: 5342次元 → 隠れ層1:16次元 → 隠れ層2:64次元
- 深層化により1層目の計算を1/3に削減 (高速・省メモリ)
 - $5342 \times 48 \times 8 \rightarrow 5342 \times 16 \times 8 + 16 \times 64 \times 8$

| | | [Tsuboi, 2014] | 提案手法 |
|-------------------|-----|----------------|---------------|
| 隠れ変数の数 | | 48 | 16, 64 |
| 活性化関数のグループ数 | | 8 | 8, 8 |
| ドロップアウト確率 | | 0.0, 0.0 | 0.2, 0.0, 0.5 |
| 開発セット | 全単語 | 97.52 | 97.52 |
| 精度 | 未知語 | 90.91 | 91.07 |
| テストセット | 全単語 | 97.51 | 97.51 |
| 精度 | 未知語 | 91.64 | 91.45 |
| ニューラルネットワークパラメータ数 | | 205万 | 69万 |
| 解析時間 | | 1 | 0.25 |

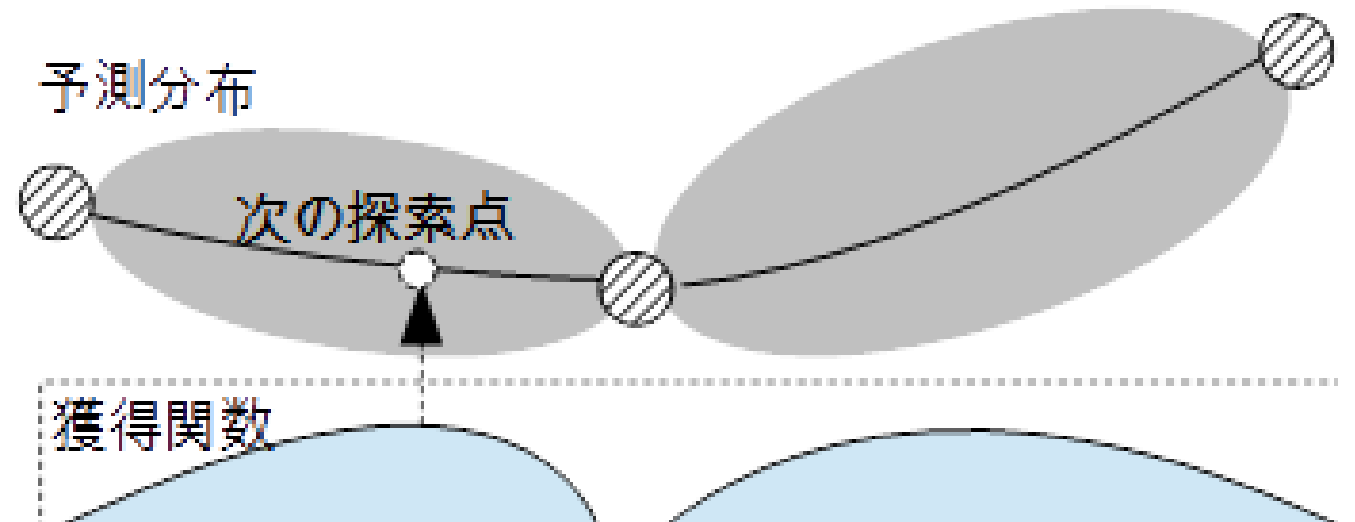
今後の課題と宣伝

• 課題

- 深層化→ハイパーパラメータ数増加→チューニング大変

- ベイズ的最適化

[Snoek et al., 2012]など
によるチューニング
自動化



- サーベイ論文: 坪井 祐太, 自然言語処理におけるディープラーニングの発展, オペレーションズ・リサーチ, Vol.60, No.4 (In press)

- 自然言語処理における課題

- リカレントニューラルネットワーク

- 畳み込みニューラルネットワーク

- 再帰ニューラルネットワーク

- フィードフォワードニューラルネットワーク