

# 先読みを用いた系列ラベリングにおける最易優先探索の適用

佐野 峻平\* 三輪 誠\*\* 鶴岡 慶雅\* 近山 隆\*  
 \*東京大学 \*\*マンチェスター大学

## 背景と動機

■ 系列ラベリング：系列内の各要素にラベル付けする問題  
 POS-tagging, チャンキング, NER

■ 履歴ベース vs CRF

	CRF	履歴ベース
最適化するもの	系列全体	系列内の各要素
ラベルの偏りの問題	あまりない	ある
精度	高	低
計算コスト	大	小
使えるラベル情報	前数語程度	どこでも

■ 履歴ベースに先読みを適用した手法 [1] が提案された  
 履歴ベースの利点を持ち、CRFと同程度の計算量と精度  
 ラベル付けは「前から後ろ」か「後ろから前」

ラベル付けを誤ると  
 後のラベル付けに  
 連鎖的に影響する

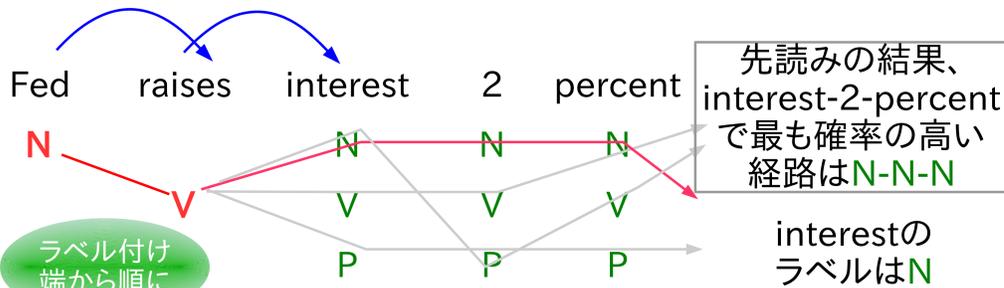
ラベル付け順序を変えたら  
 手法を改善できるかも

## 先行研究

### 先読み [1]

先読み：対象となる語より後にラベル付けされる語のラベルを決めてみて、対象となる語のラベル付け確率を計算  
 ラベル付けされていない語のラベル情報も利用できる

1. d語先までの全てのラベルの組合せについて先読みを行う
2. 確率が最も高い経路のラベル情報も用いて自身をラベル付け  
 d：先読みの深さ



### 最易優先 [2] [3]

最易：ラベル付け確率が最も高い (語, ラベル) の組  
 最易なものからラベル付けし、後のラベル付けに利用  
 早い段階でラベル付けを誤る恐れが低い

1. 全ての (語, ラベル) の組で確率を計算
2. 最易なものをラベル付け
3. ラベル付けした結果を素性に加え、周辺語の確率を更新

ラベル付け簡単な順に 最易な (語, ラベル) の組を見つけラベル付け

	Fed	raises	interest	2	percent
N	0.3	0.3	0.4	0.9	0.7
V	0.6	0.6	0.5	0.05	0.2
P	0.2	0.1	0.1	0.05	0.1

周辺の確率を更新

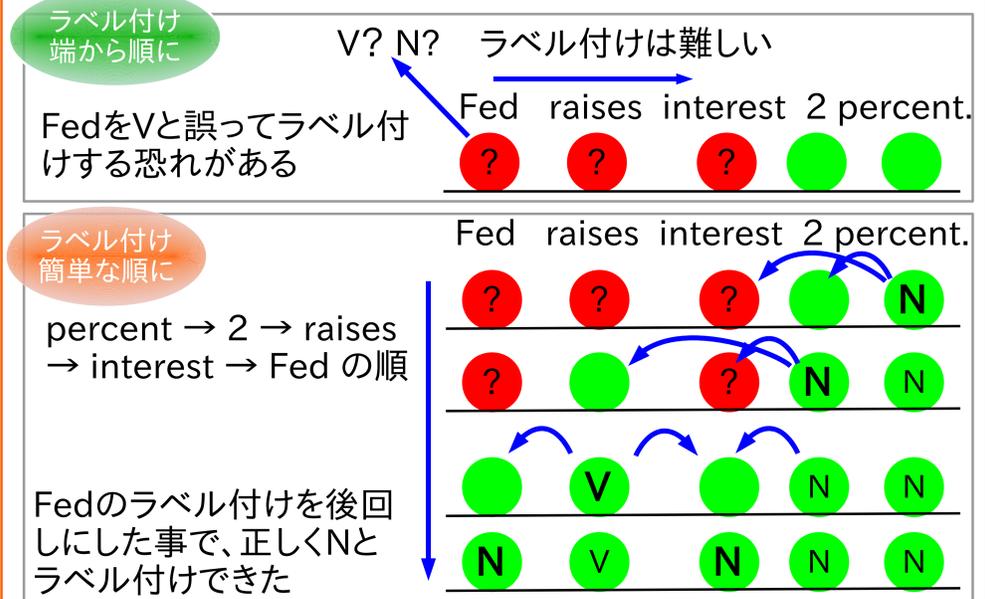
	Fed	raises	interest	2	percent
N	0.3	0.3	0.8	N	0.75
V	0.6	0.6	0.15		0.24
P	0.1	0.1	0.05		0.01

## 提案手法

### 近傍読み + 最易優先

近傍読み：対象とする語の周辺の語のラベルを決めてみる  
 ラベル付け確率の計算に近傍読みを利用し最易優先

「Fed raises interest 2 percent .」のラベル付け



●：ラベル付けの難しい語 ●：ラベル付けの容易な語

1. 全ての (語, ラベル) の組で確率を計算  
 各(語, ラベル)の組につき周辺d語のラベルを決めて素性に加えてみる  
 スコアが最も高い経路の素性を加えラベル付け確率を計算
2. 最易なものをラベル付け
3. ラベル付けした結果を素性に加え、周辺語の確率を更新  
 d：周辺読みの深さ

Fed raises interest 2 percent. における (interest, N) のスコア

raises	2	(interest, N)のスコア
N	N	10
N	V	40
N	P	15
V	N	100
P	P	5

(raises, V)(2, N) として素性を追加し(interest, N)のラベル付け確率の計算に用いる

## 課題

- 先読み (端からラベル付け)、近傍読み (最易優先でラベル付け) を実装して評価を行う
- 最易な (語, ラベル) の組を見つける際に先読みを行う手法を実装して評価を行う
- 複数の系列ラベリング問題、及び統計的係り受け解析で評価を行う

参考文献

[1] Yoshimasa Tsuruoka, Yusuke Miyao, and Jun'ichi Kazama. 2011. Learning with Lookahead: Can History-Based Models Rival Globally Optimized Models? In Proceedings of CoNLL 2011, pages 238-246.

[2] Yoshimasa Tsuruoka and Jun'ichi Tsujii. 2005. Bidirectional inference with the easiest-first strategy for tagging sequence data. In Proceedings of HLT/EMNLP 2005, pages 467-474.

[3] Yoav Goldberg and Michael Elhadad. 2010. An efficient algorithm for easy-first non-directional dependency parsing. In Proceedings of NAACL-HLT, pages 742-750.