有限状態トランスデューサ (FST) を用いた音声言語処理

塚田 元

tsukada@cslab.kecl.ntt.co.jp

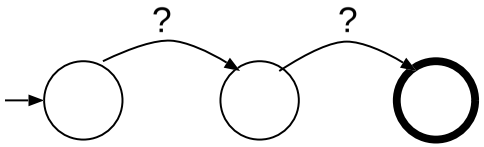
NTT コミュニケーション科学基礎研究所

2002 年 12 月 20 日

あらすじ

- 重み付き有限状態トランスデューサ (Weighted Finite State Transducer) の定式化
- WFST の基本演算
- 音声認識の定式化

種々の有限状態機械



シンボル	有限状態オートマトン (Finite State Automaton)
シンボル＋遷移確率	確率的 FSA (Probabilistic FSA) 決定的 PFSA = Markov Model 非決定的 PFSA = Hidden Markov Model
入力・出力シンボル	有限状態トランスデューサ (Finite State Transducer)
入力・出力シンボル ＋重み	重み付き FST(Weighted FST)

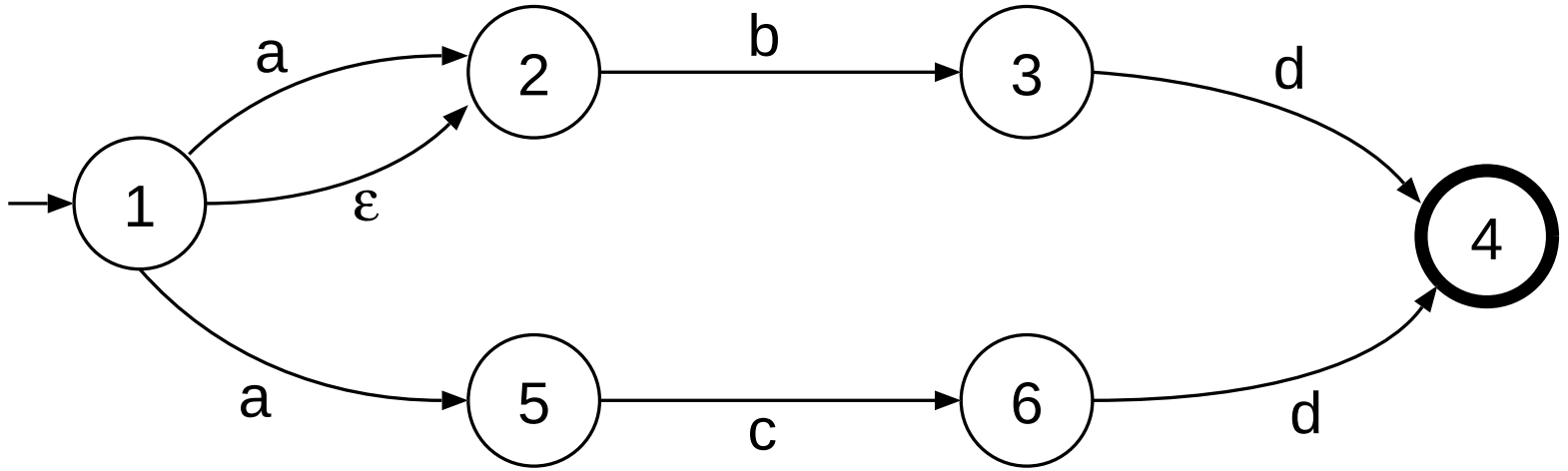
有限状態オートマトン (FSA) – 定義 –

$A = (\Sigma, Q, E, I, F)$

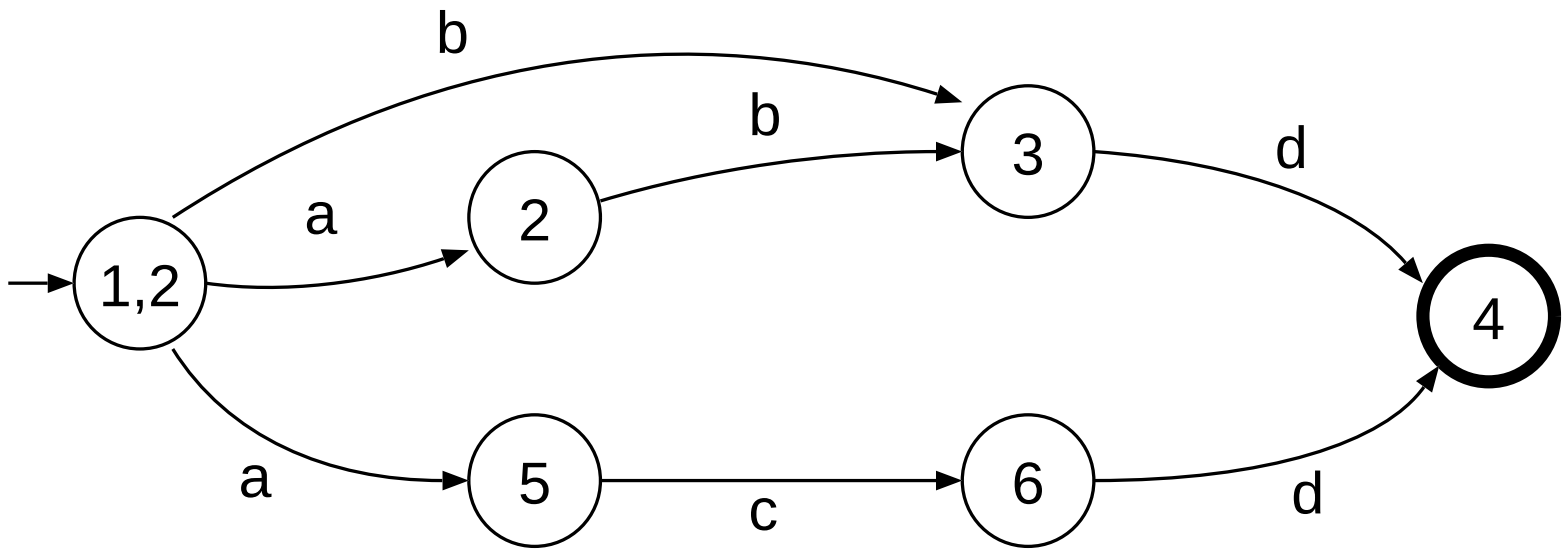
- Σ: アルファベット
- Q: 状態の有限集合
- $E : Q \times (\Sigma \cup \{\epsilon\}) \times Q$: 遷移の有限集合
- $I \subseteq Q$: 初期状態の集合
- $F \subseteq Q$: 最終状態の集合

全ての FSA は決定的でかつ状態数最小の等価な FSA に変換可能.

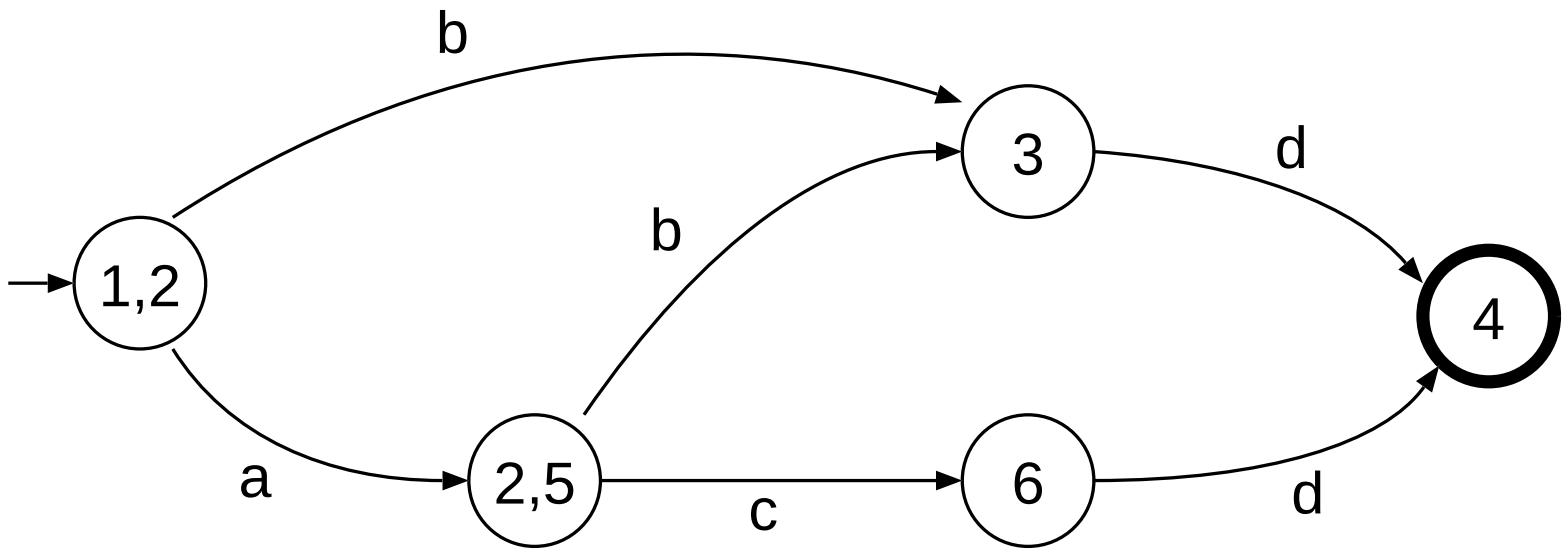
非決定的 FSA の例



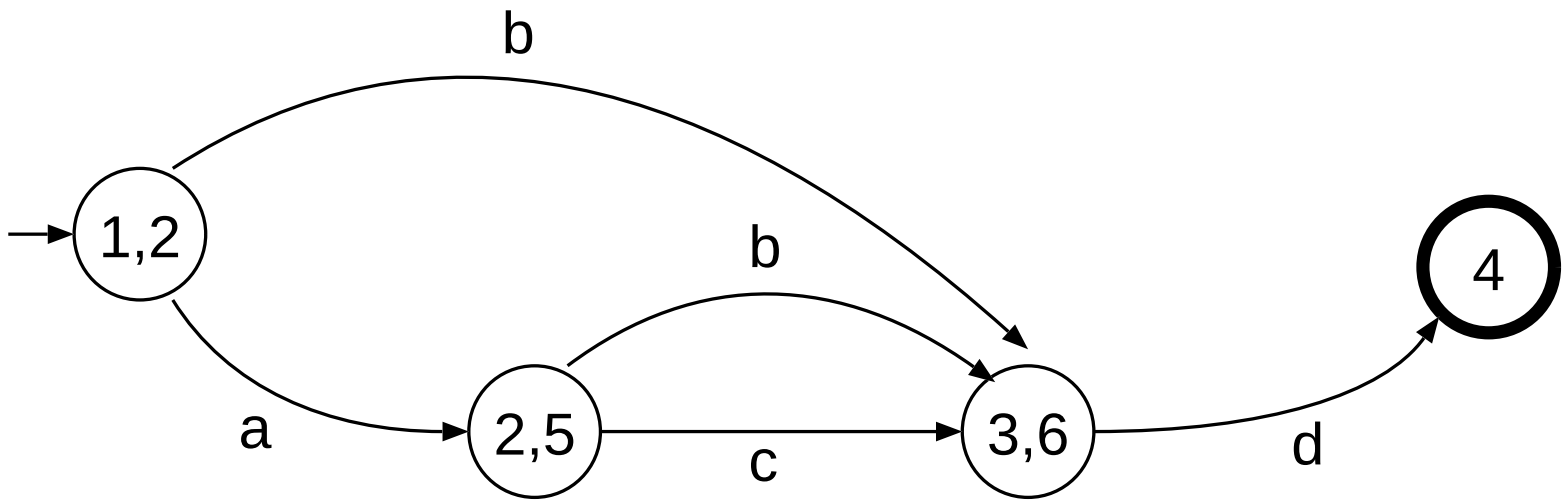
ϵ の消去結果



決定化の結果



最小化の結果



重み付き有限状態トランスデューサ (WFST) –定義–

$A = (\Sigma, \Delta, Q, K, E, I, F)$

- Σ: 入力アルファベット
- Δ: 出力アルファベット
- Q: 状態の有限集合
- K: 重みの半環
- $E : Q \times (\Sigma \cup \{\epsilon\}) \times (\Delta \cup \{\epsilon\}) \times K \times Q$: 遷移の有限集合
- $\lambda : I \to K$: 初期状態重み関数
- $\rho : F \to K$： 最終状態重み関数
- $I \subseteq Q$: 初期状態の集合
- $F \subseteq Q$: 最終状態の集合

半環 (semiring)

半環 $(K, \oplus, \otimes, \overline{0}, \overline{1})$: 逆元なしの環 (ring)

半環の種類	集合	$\oplus$	$\otimes$	$\overline{0}$	$\overline{1}$
確率値	$[0, 1]$	$+$	$\times$	0	1
log 確率値 (Viterbi 近似)	$[-\infty, 0]$	max	$+$	$-\infty$	0
tropical	$[-\infty, +\infty]$	min	$+$	$+\infty$	0
string	$\Sigma^* \cup \{\infty\}$	longest common prefix	連結	∞	ϵ

演算: 合成 (composition)

入力

二つのトランスデューサ T_1, T_2

出力

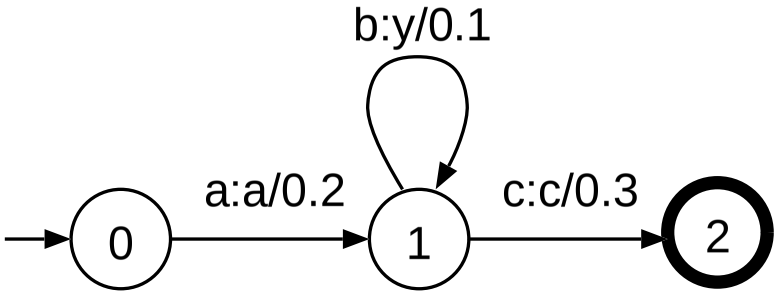
以下の式を満たす $T_1 \circ T_2$

$$[T_1 \circ T_2](x, y) = \bigoplus_z [T_1](x, z) \otimes [T_2](z, y)$$

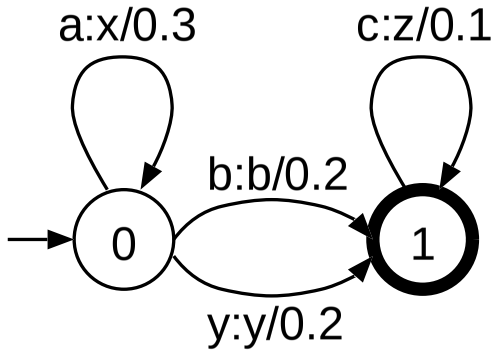
演算: 合成 (composition)

例: $\otimes$ が $+$ の半環

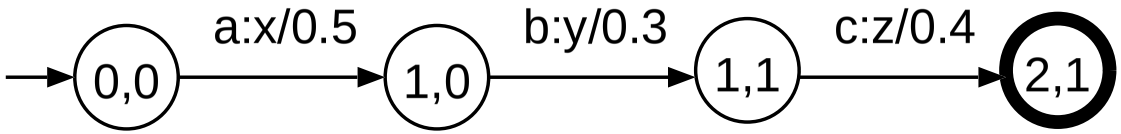
T_1



T_2



$T_1 \circ T_2$

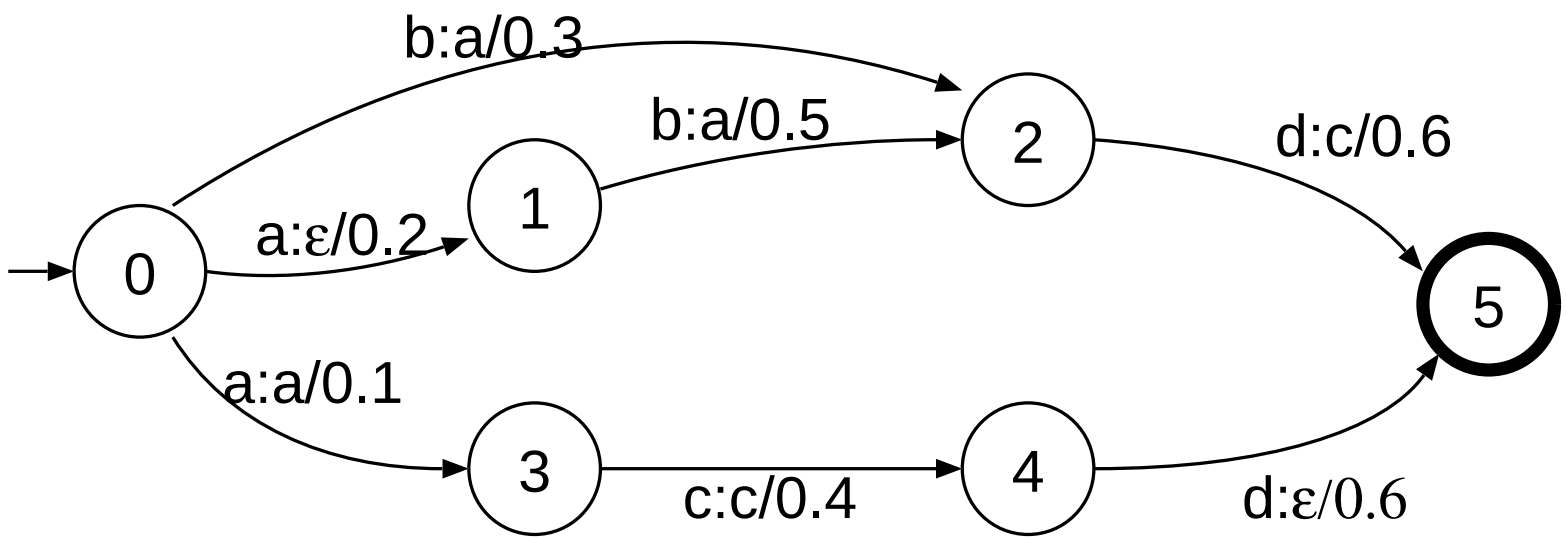


演算: 決定化 (determinization)

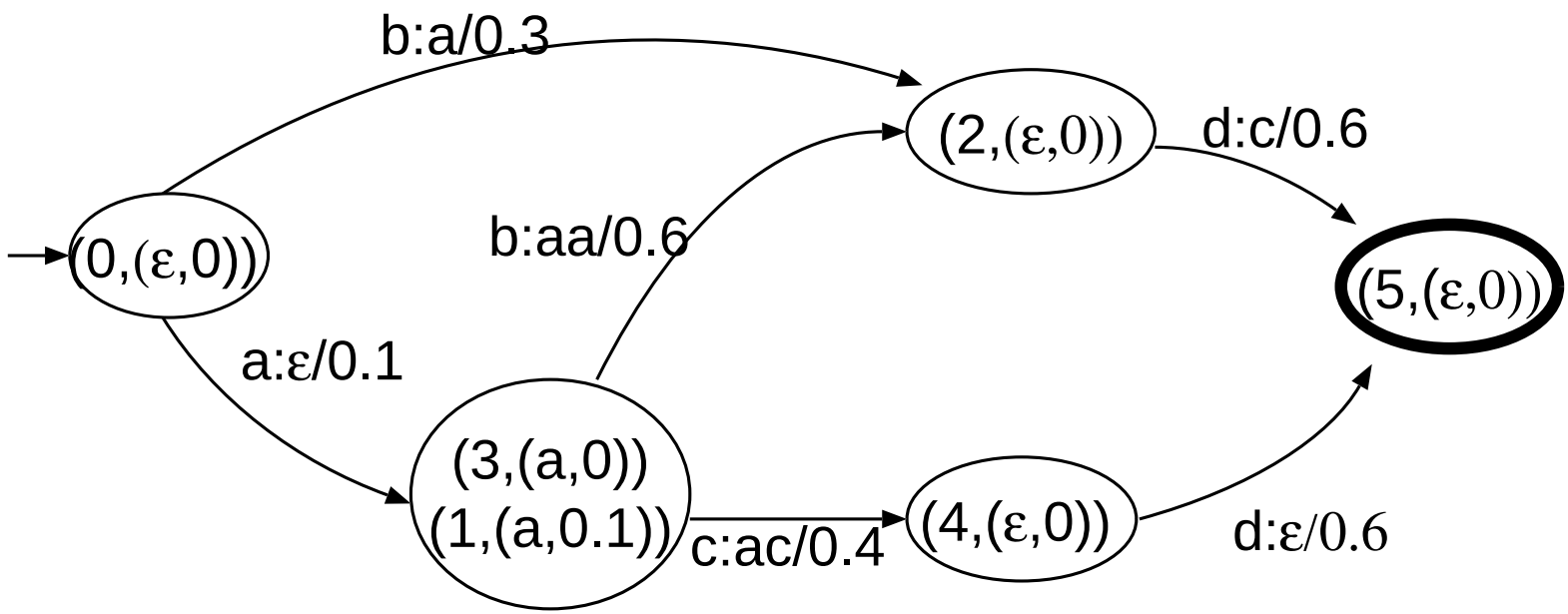
入力： WFST
出力： １つの初期状態をもち、１つの入力アルファベットに対して、
１つの状態から高々１つの遷移をもつ WFST。

- 半環の条件： weekly left divisible
- 必ずしも全ての WFST が決定化できるわけではない。
- 非循環は決定化可能であることの十分条件。

決定化の入力



決定化の出力



状態： (元の状態, (未出力記号, 未出力重み)) の集合

出力関数： $\delta'(S,a) = \bigoplus_{(q,u)\in S} \bigoplus_{q'\in Trans(q,a)} u \otimes \delta(q,a,q')$

遷移関数：

$Trans'(S,a) = \bigcup_{(q,u)\in S} \bigcup_{q'\in Trans(q,a)} \{(q', \delta'(S,a)^{-1}(u \otimes \delta(q,a,q')))\}$

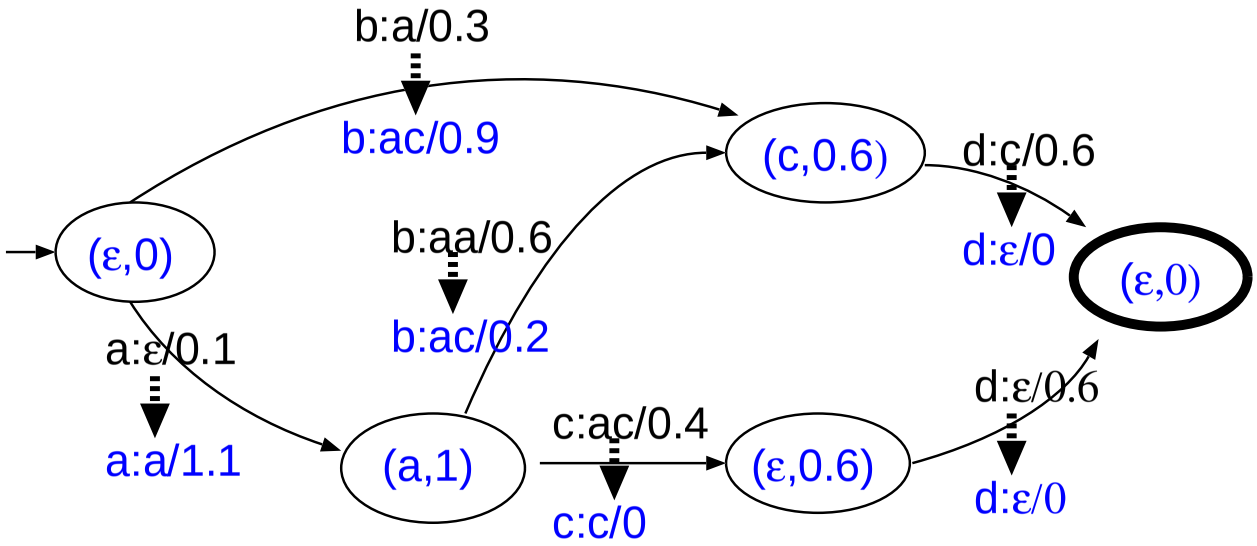
演算： pushing

入力： WFST

出力： 可能な限り出力アルファベットを早期に出力し、重みについても極力初期状態に引き寄せた WFST。(或る種の正規化手法)

- 条件： weekly divisible semiring, zero-sum free semiring/machine

pushing の例



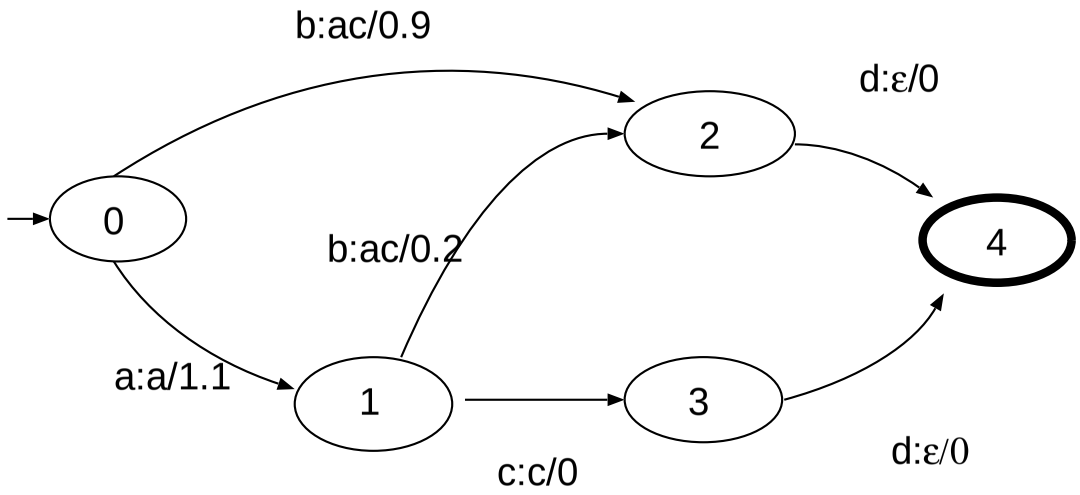
- 最終状態まで至る全てのパスにおいて、出力列の『(longest common prefix, 出力重みの和)』を、各状態に保持。
- 各状態の値： $distance(q) = \bigoplus_{\pi \in Path(q,Fin)} out(\pi)$
- 出力関数：
  $out'(e) = distance(src(e))^{-1}(out(e) \otimes distance(dest(e)))$

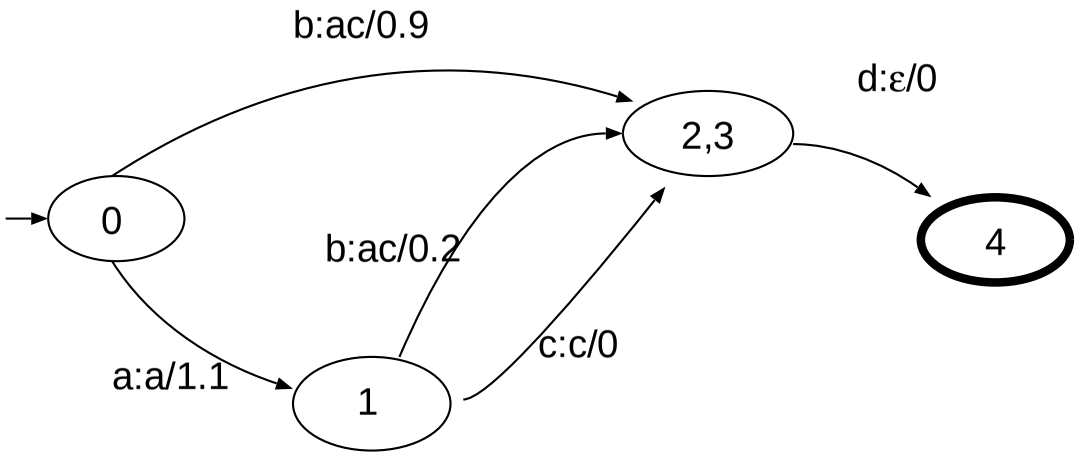
演算： 最小化 (minimization)

入力： 決定的な WFST
出力： 状態数が最小の決定的 WFST

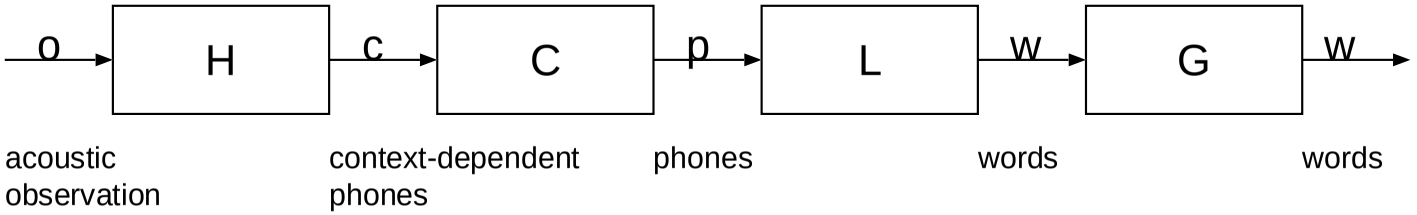
- pushing により正規化
- 通常の FSA の最小化アルゴリズムを適用

最小化の例



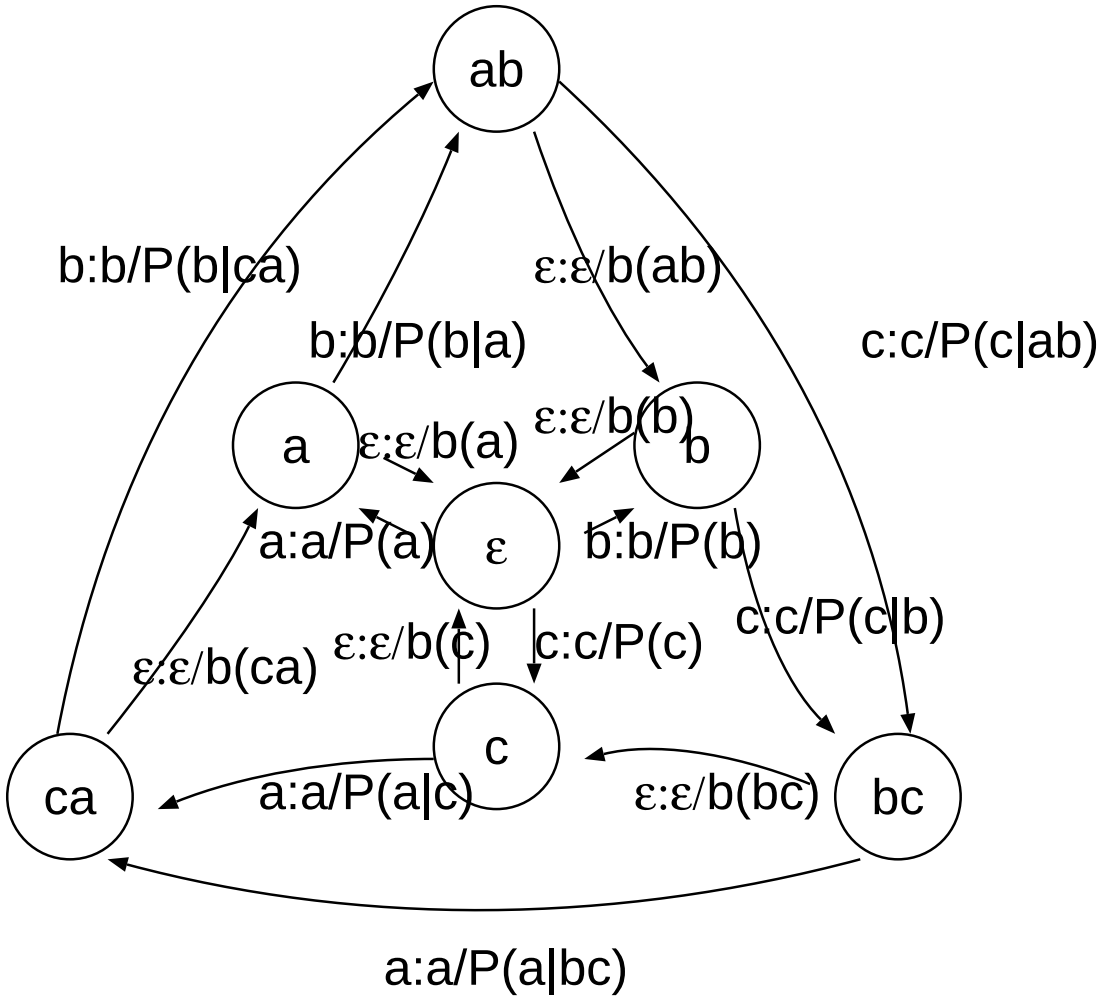


音声認識のモデル化

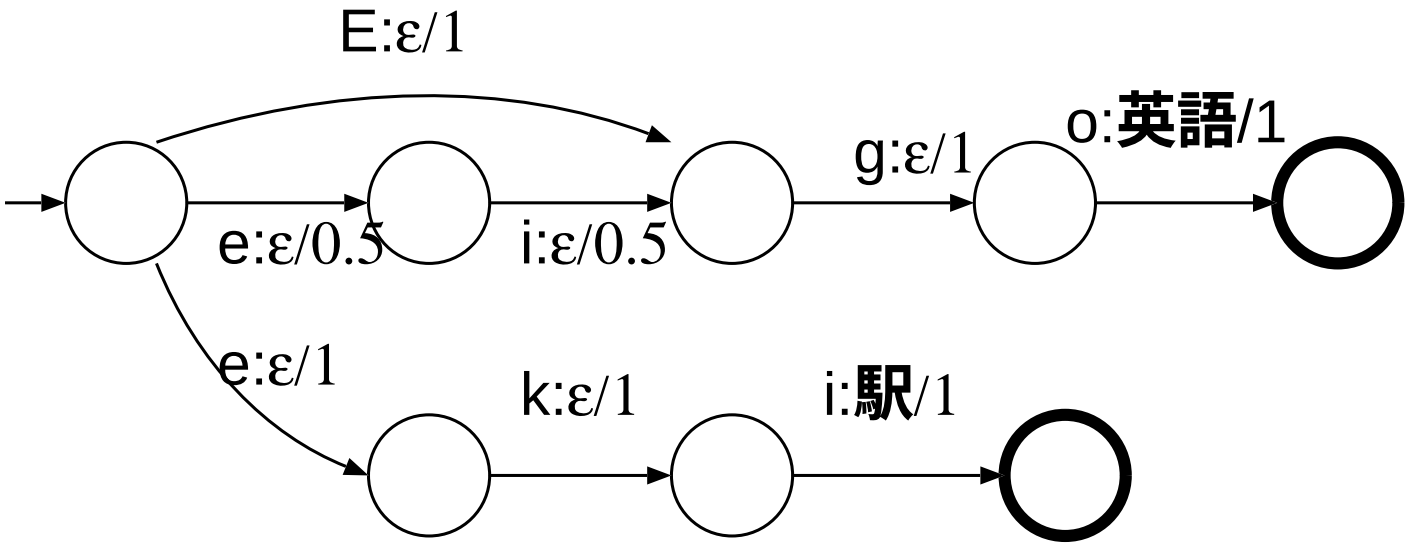


- $argmax_w P(w|o) = argmax_w P(o|w)P(w) = argmax_w P(o|c)P(c|p)P(p|w)P(w)$
- 各々の $P(x|y)$ を WFST によりモデル化。o を入力としたとき $H \circ C \circ L \circ G$ が出力する最も尤もらしい w を求める問題として定式化。
- メモリと計算時間のトレードオフにより、静的に合成 (composition) しておく範囲を選択できる。
- 木構造辞書により音響・言語尤度を共有化する技術の一般化とみなせる。静的に合成しておく範囲を増やすことで、サーチエラーを減らすことができる。

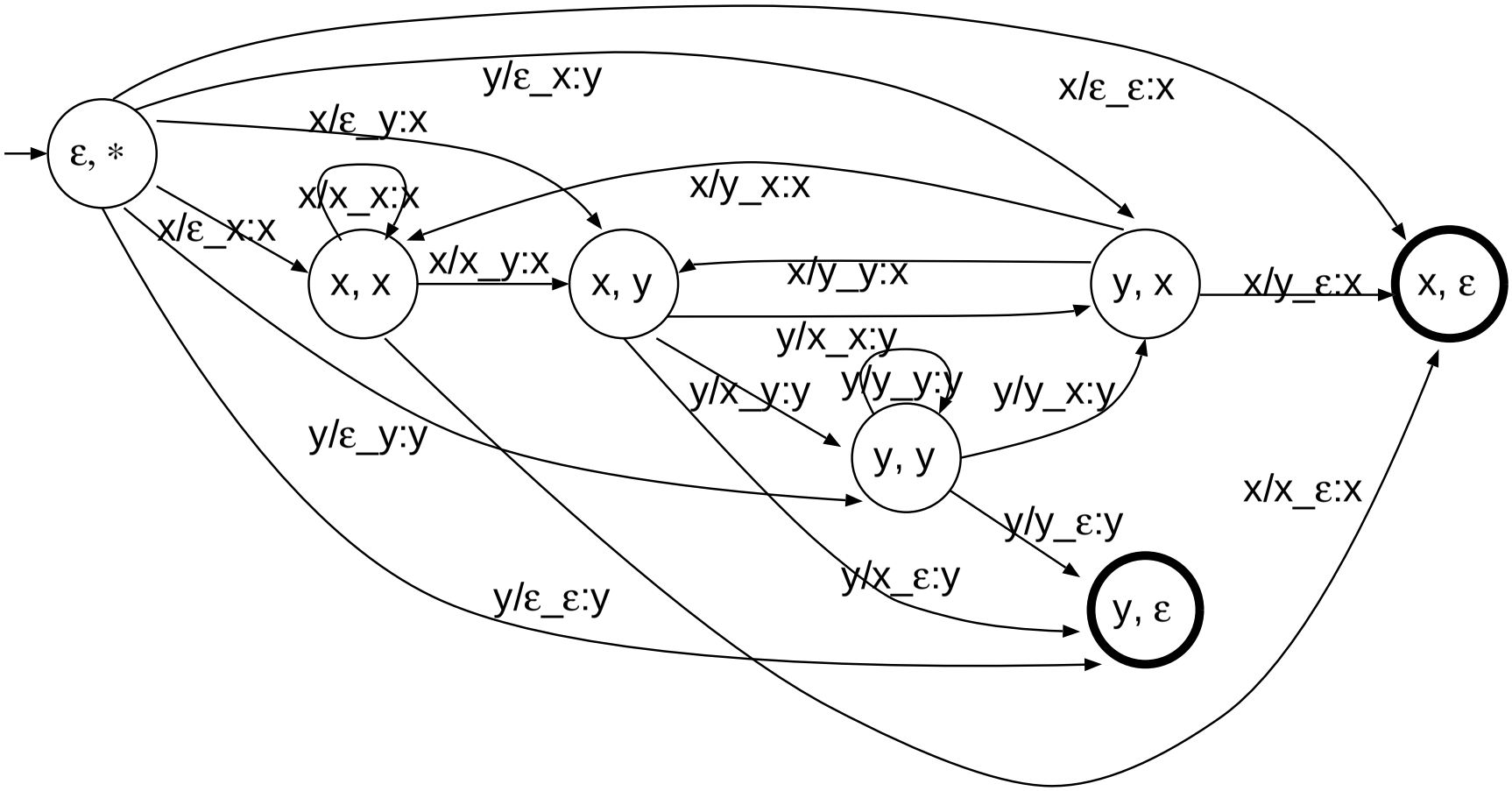
trigram の例



発音辞書の例



Triphone トランスデューサの例



実用上の工夫

- 言語モデルのネットワークサイズを爆発させないために、ϵ を普通の記号として扱う。
- 決定化が可能になるように、仮想的な音素を導入し同音異義語を無くす。

  k i t a #1　北

  k i t a #2　来た

  この仮想記号を出力するような自己ループを H や C に付け加える。

講演音声の認識実験 – ネットワークサイズ –

ネットワーク	状態数	遷移数
H	33,767	50,672
C	795	39,210
L	129,370	149,693
G	159,427	517,597
$H \circ C \circ L \circ G$	4,413,443	9,264,432
$H \circ min(det(C \circ L \circ G))$	2,068,728	5,925,541
$min(det(H \circ min(det(C \circ L \circ G))))$	1,742,876	4,618,082

※ min は pushing を含む。

講演音声の認識実験 – サーチエラーと探索効率 –

- 認識対象： 日本語話し言葉コーパス テストセット (4 講演)
- 認識結果： 単語誤り率 (処理時間 RTF)

ネットワーク	尤度幅 90		尤度幅 100	
$H \circ C \circ L \circ G$	51.8	(17.5)	51.2	(21.8)
$H \circ min(det(C \circ L \circ G))$	38.5	(3.0)	36.5	(4.1)
$min(det(H \circ min(det(C \circ L \circ G))))$	36.8	(2.2)	35.2	(3.0)

音声認識実験は、Daniel Willett 氏の作成したデコーダ、MIT の FST ツールを用いて、堀 貴明氏に実施いただいた。

まとめ

- 音響モデルから言語モデルまでを融合したモデルがひとつのWFST として実現可能になり、大語彙連続音声認識に適用されるようになってきた。
- 従来の探索手法に比べ、サーチエラーを減らすことが可能。

今後の研究課題

- ネットワークサイズの縮小化
- 本手法と相性のよい近似手法
- パラメータおよび構造の学習
- 全体最適化
- より高度な言語処理 (トピック抽出、要約、機械翻訳など) との融合

参考文献

基本演算

Roche and Schabes eds, Finite-State Language Processing, The MIT Press, 1997.

Mohri, Finite-State Transducers in Language and Speech Processing, Computational Linguistics, 23(2), 1997.

Mohri, Minimization Algorithms for Sequential Transducers, Theoretical Computer Science, 234, 2000.

大語彙連続音声認識

Mohri, Pereira, and Riley, Weighted Finite-State Transducers in Speech Recognition. Computer Speech and Language, 16(1), 2002.

チュートリアル

Mohri and Riley, Weighted Finite-State Transducers in Speech Recognition, ICSLP'02