

Buried Markov Model を用いた 音声認識モデルの構築法の検討

山本 隆之^{†1} 滝口 哲也^{†2} 有木 康雄^{†2}

HMM による音声認識は、はっきりと発話された音声に対しては高い精度で認識を行えるが、雑音を含む音声や、連続的に発話された音声に対して精度が低下するという問題がある。この問題に対し、観測ノード間の時間的な依存関係を記述できる Buried Markov Model を用いた音声認識モデルが J. Bilmes によって提案された。本研究ではこの Buried Markov Model の時間的依存関係の学習方法である Pairwise アルゴリズムにおいて従来用いられている相互情報量の代わりに、ノンパラメトリックな依存性の検定法を導入することで、分布を仮定した場合には得られない依存関係を取得し、モデルを構築することで認識精度の改善を試みた。音素認識精度の結果、相互情報量を用いた場合に比べ、ノンパラメトリックな独立性検定法である多重解像度独立性検定を用いた場合、男性話者で 0.6 %、女性話者で 0.6 %、混合話者で 1.3 % の音素認識精度の改善が確認された。

Structuring Model for Speech Recognition using Buried Markov Model

TAKAYUKI YAMAMOTO,^{†1} TETSUYA TAKIGUCHI ^{†2}
and YASUO ARIKI^{†2}

Though HMM makes it possible to recognize clear utterance by high accuracy, there is a problem that the speech including the noise or the continuous utterance cause the decrease in accuracy. To solve this problem, buried Markov model was proposed by J. Bilmes, where Buried Markov model contains the conditional independence between the observation node. In this research, to enable to obtain the conditional independence that the assumption of distribution prevents from obtaining, we proposed to introduce non-parametric independence test instead of mutual information into pairwise algorithm, that is the algorithm to learn the structure of buried Markov model. We performed phone recognition experiment by HMM, BMM structured using mutual information and BMM structured using multi-resolution independence test. Experimental results achieved improvement of phone recognition rate of male of 0.6%, im-

provement of phone recognition rate of female of 0.6%, and improvement of phone recognition rate of mix of 1.3%.

1. はじめに

現在、音声認識において、音響モデルとして隠れマルコフモデル (Hidden Markov Model, HMM) が広く用いられている。HMM はダイナミックベイジアンネットワーク (Dynamic Bayesian Network, DBN) の一種で、隠れ状態と観測のみからなる単純な構造の確率モデルである。HMM のこういった単純な構造は、2つの独立仮定の上に成り立っている。1つは“状態は1フレーム前の状態のみによって決まる”という仮定であり、もう1つは“観測はフレームごとに独立であり各フレームの状態のみによって決まる”という仮定である。この2つの仮定によって音声の生成構造を単純化しているため、モデル自体は扱いやすく、明瞭な発音の音声に対しては高い精度で認識を行うことが可能であるが、複雑な事象を表現するには適しておらず、ノイズが含まれる音声や会話などの話し言葉の認識において、性能が著しく低下するという問題がある。

このような問題に対して、前述の HMM の仮定を緩和し、HMM の表現能力を高めることでより複雑な事象に対応できるモデルを作成し、認識精度の向上を図るという試みがなされている。J. Bilmes によって考案された Buried Markov Model (BMM)¹⁾ もそういった試みの1つであるといえる。BMM は HMM の隠れ状態と観測の構造に加え、現在の観測と他フレームの観測間の依存関係及び現在の観測の内部での依存関係を表現できる確率モデルである。この BMM の依存関係はデータを学習することで求められるが、この学習法については十分に検討がなされていない。本研究では、BMM の構造の学習に必要な独立性の検定法について複数の手法を導入し、従来用いられている相互情報量を用いて学習した BMM との連続音声の認識精度を比較し、より識別性能の高い認識モデルを構築できる独立性検定法について検討を行った。

以降の2章では BMM の概要とその構造学習法を解説し、3章では BMM の構築に用いる独立性検定について紹介する。4章に実験の概要及び考察を示し、5章に本稿のまとめを

^{†1} 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

^{†2} 神戸大学自然科学系先端融合研究環

Organization of Advanced Science and Technology, Kobe University

述べる。

2. Buried Markov Model

2.1 概 要

Buried Markov Model は Hidden Markov Model の各フレームの観測ノード間に依存関係を加えた図 1 のような構造のグラフィカルモデルによって表わされる。HMM では観測可能な情報のみが観測され、マルコフ過程に従う各フレームの状態が実際に観測されない隠れた状態であることから、隠れ (hidden) マルコフモデルと呼ばれるが、BMM ではこの隠れた状態のマルコフ過程が観測ノード間の依存関係によってより隠れた状態 (buried) になっていることから、Buried Markov Model と呼ばれている。

BMM の観測ノード間の依存関係の構造は、参照するフレームの隠れ状態によって決まるため、各隠れ状態ごとに学習が行われる。依存関係の構造はある状態における参照フレームのノードと過去や未来のフレームのノードの間の相関関係を調べることで学習される。こ

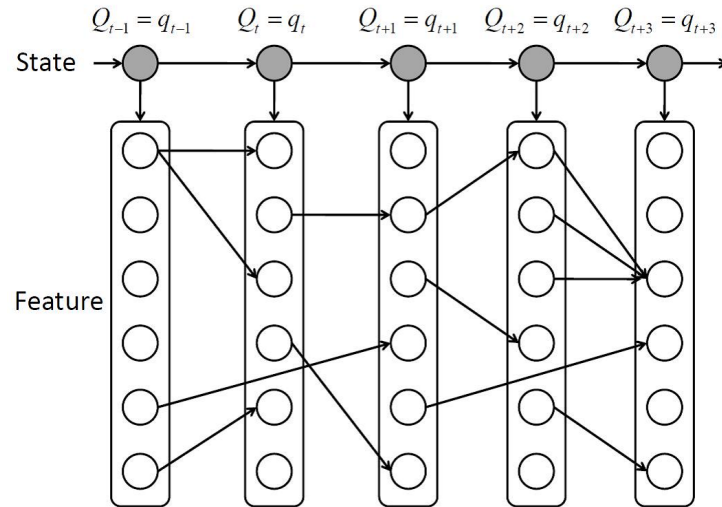


図 1 Buried Markov Model の構造
Fig.1 Structure of Buried Markov Model

の依存関係の学習アルゴリズムについては次節で述べる。

また BMM を式の形で表すと式 (1) のような形で表される。

$$Pr(x_{1:T}) = \prod_t Pr(x_t | z_t(q_t), q_t) Pr(q_t | q_{t-1}) \quad (1)$$

$x_{1:T}$ はフレーム数 T のデータからの観測を表し、 q は観測に対応する状態、 z_t はフレーム t における観測間のエッジの集合を表す。この z_t はフレーム t における状態 q_t によって一意に決まる関数である。

2.2 構造学習

BMM の構造学習には相互情報量を用いた Pairwise アルゴリズムが用いられる。Pairwise アルゴリズムはターゲットノードに対し、相互情報量が次の 3 つの条件を満たすノードを親ノードと判定する方法である。

$$I(X, Z | q) > \delta_1 \quad (2)$$

$$I(Z, Z_i) < \delta_2 \quad I(X, Z) \quad Z_i \in \mathbf{Z} \quad (3)$$

$$I(X, Z) < \delta_3 \quad (4)$$

X はターゲットノード、 Z は親の候補、 $\mathbf{Z}$ はすでに親であることが決定しているノードの集合、 $\delta_1, \dots, \delta_3$ はそれぞれの閾値を表す。

まず式 (2) は親ノードの候補を相関の高いノードに絞るという条件である。ターゲットノード X とは別フレームのノード Z でも、状態 q の特徴を表しているノードである X との相関が高ければ、状態 q の特徴を表す情報として有効であると考え、閾値を設定し、相互情報量がこの閾値よりも高いノードを親の候補とする。しかし、相関の高いノードを親としても、他の状態においても同様に相関が高ければ、状態 q の特徴を良く表したノードであるとは言い切れない。そこで式 (4) の条件を導入し、 X, Z の各状態での相互情報量の平均が設定した閾値より小さいかどうか判定する。状態 q での相互情報量が高く、他の状態での相互情報量が小さければ、状態 q の特徴を良く表したノードであると考え、親の候補とする。式 (3) は条件付き独立性の検定を行っている。ターゲットノード X と候補ノード Z との相関が高いとしても、 Z と X の親ノード Z_i との相関が強ければ、 Z と X の関係は Z_i と X の関係と同種のものとなり、 Z を親とすることは冗長であると考えられるため、 Z を候補から外す。

また関数 $U(Z) = I(X, Z | q) - I(X, Z)$ を用いて候補ノード Z がターゲット X に対して持っている情報の大きさを計算し、この値が大きいものから順に前述の 3 つの条件を調べ、状態ごとの親ノードセットを出力として得る。BMM の学習は観測間の依存関係を

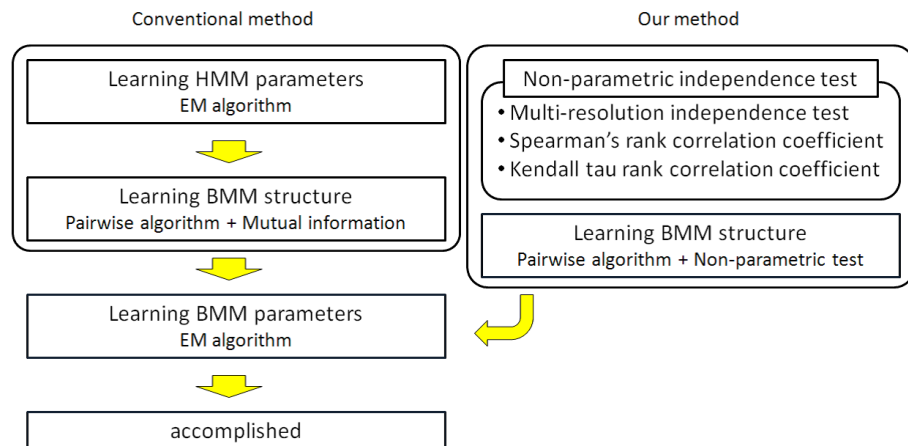


図2 Buried Markov Model の構築法
Fig. 2 Method to structure Buried Markov Model

決定していない状態の BMM，すなわち HMM の状態で HMM パラメータの学習を行い，Pairwise アルゴリズムを用いて BMM の構造を決定し，観測間のエッジが決定された状態で再度パラメータを学習するという手順で行う。この手順をとるのは Pairwise アルゴリズムで用いる相互情報量を計算するにあたって，初期学習した HMM のパラメータを用いるためである。この方法では HMM のモデリング精度によって得られる依存関係が変わってくる。これを避けるため相互情報量の代わりに次章で述べるノンパラメトリック検定を導入する。

3. 独立性検定

本章では BMM の構築に用いる独立性の検定法について紹介する。従来 BMM の構築に用いられる相互情報量はパラメトリックに独立性を検定する方法であり，分布を限定することで認識に有効な相関の情報が欠落してしまっている可能性がある。そこで独立性の検定手法として，多重解像度独立性検定²⁾，スピアマンの順位相関係数³⁾，ケンドールの順位相関係数⁴⁾といったノンパラメトリックな検定法を導入し，パラメトリックな独立性検定では欠落している相関の情報を取り入れて認識モデルを構築することで，認識精度の改善を試

みた。

3.1 多重解像度独立性検定

多重解像度独立性検定 (Multi-resolution Independence Test) は 2 つの確率変数をプロットした散布図をある解像度で分割し，作成した 2 次元ヒストグラムの情報からデータの大域的な変動の情報を取り出し，確率変数間の独立性をノンパラメトリックに計算する検定法である。このデータ分割の解像度を上げていき，データの発生尤度が最大となる分割を発見し，その時の尤度を用いて独立性を計算することから，多重解像度独立性検定と呼ばれている。

はじめに，解像度と境界を固定した場合の独立性検定を考える。まず，解像度を $R \equiv I \times J$ とし，2 つの確率変数のサンプルの存在区間を $I \times J$ の領域に分割する (図 3)。各領域に含まれるサンプルの数をそれぞれ $c_1, \dots, c_K$ ， $K \equiv IJ$ ，サンプルの総数を N ，各領域でサンプルが発生する確率を $p_1, \dots, p_K$ ，各領域の境界の集合を $\mathbf{B}_R$ とすると 2 つの確率変数のサンプルからなるデータ $\mathbf{D}$ の尤度は次式で表される。

$$Pr(\mathbf{D} | p_{1 \dots K}, \mathbf{B}_R, R) = N! \prod_{k=1}^K \frac{p_k^{c_k}}{c_k!} \quad (5)$$

式 (5) においてパラメータ $p_{1 \dots K}$ は与えられていないため，次式で表される多項分布の共役事前分布であるディリクレ分布を事前分布として与える。

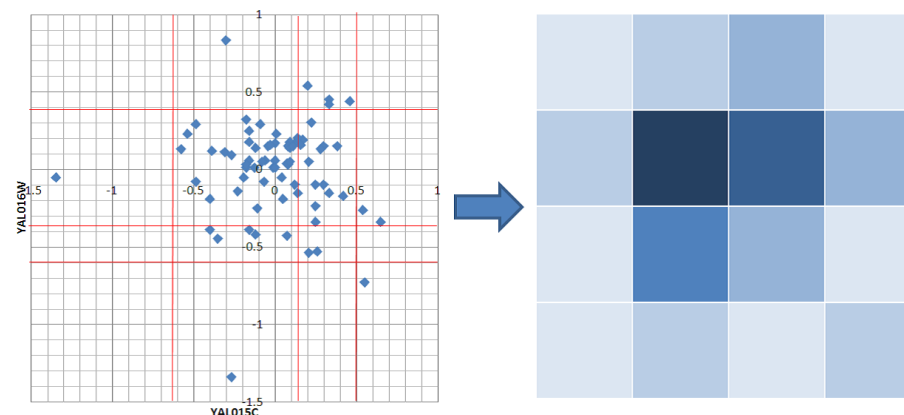


図3 ヒストグラムの作成
Fig. 3 Histogramization

$$Pr(p_{1...K}) = \Gamma(\gamma) \prod_{k=1}^K \frac{p_k^{\gamma_k-1}}{\Gamma(\gamma_k)} \quad (6)$$

式 (6) において $\gamma = \sum_{k=1}^K \gamma_k$ であり, γ_k はディリクレ分布の設定パラメータ, $\Gamma(x)$ はガンマ関数を表す. また, データ尤度 $Pr(\mathbf{D})$ を p_k で周辺化し, 式 (5), (6) を代入すると, 固定解像度・固定境界におけるデータ尤度 $Pr(\mathbf{D})$ が得られる.

$$\begin{aligned} Pr(\mathbf{D}) &= \int Pr(\mathbf{D}|p_{1...K}) Pr(p_{1...K}) dp_{1...K} \\ &= \frac{\Gamma(\gamma)}{\Gamma(\gamma+N)} \prod_{k=1}^K \frac{\Gamma(\gamma_k + c_k)}{\Gamma(\gamma_k)} \end{aligned} \quad (7)$$

データ $\mathbf{D}$ が独立モデル M_I , 依存モデル M_{-I} のどちらか一方から生成されると仮定すると, これについて周辺化した式は次式の形になる.

$$Pr(\mathbf{D}) = Pr(\mathbf{D}|M_I)Pr(M_I) + Pr(\mathbf{D}|M_{-I})Pr(M_{-I}) \quad (8)$$

式 (8) をベイズの定理を用いて変形し, 両モデルの事前確率をそれぞれ $Pr(M_I) \equiv \rho$, $Pr(M_{-I}) \equiv 1 - \rho$ とすると次式が得られる.

$$Pr(M_I|\mathbf{D}) = 1 / \left[1 + \frac{1-\rho}{\rho} \frac{Pr(\mathbf{D}|M_{-I})}{Pr(\mathbf{D}|M_I)} \right] \quad (9)$$

依存モデルにおいては, 1つの多項分布を仮定すれば良いが, 独立モデルにおいては, 2つの確率変数が独立であるため両軸方向に多項分布を仮定しなければならない. よってデータ尤度 $Pr(\mathbf{D}|M_{-I})$, $Pr(\mathbf{D}|M_I)$ は式 (9) を用いて次のように表される.

$$\begin{aligned} Pr(\mathbf{D}|M_{-I}) &= \frac{\Gamma(\gamma)}{\Gamma(\gamma+N)} \prod_{k=1}^K \frac{\Gamma(\gamma_k + c_k)}{\Gamma(\gamma_k)} \\ &\equiv \Upsilon(\mathbf{C}_K, \gamma_K) \end{aligned} \quad (10)$$

$$\begin{aligned} Pr(\mathbf{D}|M_I) &= \frac{\Gamma(\alpha)}{\Gamma(\alpha+N)} \prod_{i=1}^I \frac{\Gamma(\alpha_i + c_i)}{\Gamma(\alpha_i)} \times \frac{\Gamma(\beta)}{\Gamma(\beta+N)} \prod_{j=1}^J \frac{\Gamma(\beta_j + c_j)}{\Gamma(\beta_j)} \\ &\equiv \Upsilon(\mathbf{C}_I, \alpha_I) \Upsilon(\mathbf{C}_J, \beta_J) \end{aligned} \quad (11)$$

式 (10), (11) において $\gamma = \sum_{k=1}^K \gamma_k$, $\alpha = \sum_{i=1}^I \alpha_i$, $\beta = \sum_{j=1}^J \beta_j$ とし, 事前分布を一様分布とするため, $\alpha_i = \beta_j = \gamma_k = 1$ と設定する. 式 (10), (11) を式 (9) に代入し, 次式を

得る.

$$Pr(M_I|\mathbf{D}) = 1 / \left[1 + \frac{(1-\rho)\Upsilon(\mathbf{C}_K, \gamma_K)}{\rho\Upsilon(\mathbf{C}_I, \alpha_I)\Upsilon(\mathbf{C}_J, \beta_J)} \right] \quad (12)$$

式 (12) により固定解像度・固定境界でのモデル尤度を計算することができる. これを多重解像度で計算し, 最小の独立性を求める. ある解像度 R における依存性は式 (13) で与えられる.

$$Pr(M_I|R, \mathbf{D}) = \int Pr(M_I|\mathbf{B}_R, R, \mathbf{D}) Pr(\mathbf{B}_R|R, \mathbf{D}) d\mathbf{B}_R \quad (13)$$

この式において, $\mathbf{B}_R$ は, 解像度 R における境界集合であり, 図 3 において赤線で表された境界の集合を指す. この式は解像度 R において取り得る全ての境界の組み合わせに関して $Pr(M_I|\mathbf{D})$ を計算し, これの重み付き平均を解像度 R での独立性とすることを意味する. また境界には各データ間の中点を通る直線を用い, $Pr(\mathbf{B}_R|R, \mathbf{D})$ にはデータ間の距離をデータの存在区間の長さで割った値を用いる. 解像度 R での独立性を計算した後, さらに解像度を上げ, 繰り返し独立性を計算し, 独立性が最小となる解像度 R_{max} を求め, その時点の独立性を変数間の独立性とする. しかし解像度 R において取り得る全ての境界の組み合わせは R が増加するにつれて大幅に増加するため, 境界は独立性を最小とする軸を各軸 1 本ずつ加えていき, 加えた軸に関しては除外して考える.

以上の手順で独立性を検定するのが多重解像度独立性検定である. この方法で独立性を検定することにより, 線形の相関だけでなく, 非線形の複雑な相関についても判定することが可能となる.

3.2 スピアマンの順位相関係数

スピアマンの順位相関係数 (Spearman's rank correlation coefficient) は 2 つの変数の順位の間の相関の強さを表す指標であり, 次式で表わされる.

$$\rho = 1 - \frac{6 \sum_{i=1}^n D_i^2}{(n-1)n(n+1)} \quad (14)$$

長さ n の 2 つのデータ系列 X , Y におけるあるインデックス i におけるデータ X_i, Y_i の順位の差を $D_i = \text{rank}(X_i) - \text{rank}(Y_i)$ とする. ρ はデータの順位のピアソンの相関係数を計算したものである. ピアソンの相関係数ではデータが正規分布に従うことを仮定しなければならないが, データを直接入れるのではなく, 順位を入れることで分布を仮定せずに 2 つのデータの相関性を導出することを可能としている.

3.3 ケンドールの順位相関係数

ケンドールの順位相関係数 (Kendall tau rank correlation coefficient) はスピアマンの順位相関係数と同様の2つの変数の順位の相関の強さを表す指標であり、次式で表わされる。

$$\tau = \frac{P - Q}{\frac{1}{2}n(n-1)} \quad (15)$$

2つの長さ n のデータ系列 X, Y から2つのデータを選ぶ。この選び方は $\frac{1}{2}n(n-1)$ 通り存在する。選んだデータのインデックスを a, b とすると、 P は「 $X_a < X_b$ かつ $Y_a < Y_b$ 」または「 $X_a > X_b$ かつ $Y_a > Y_b$ 」の条件に当てはまる組み合わせの数、 Q は「 $X_a < X_b$ かつ $Y_a > Y_b$ 」または「 $X_a > X_b$ かつ $Y_a < Y_b$ 」の条件に当てはまる組み合わせの数である。これはすなわち、 P が2つのデータ間の正の相関性を、 Q が負の相関性を表しており、 P が大きくなれば正の相関が強くなり τ は1に近づき、 Q が大きくなれば負の相関が強くなり τ は-1に近づく。 P と Q が同じであれば τ は0となりデータ間の相関はないと判断できる。ケンドールの順位相関係数はノンパラメトリックに相関性を検定することができる指標である。その性質上線形・非線形に関わらず相関関係を判別することができるが、複雑な相関関係を判別することができず、単調関数であるかを判別する指標となる。

4. 実 験

BMM の学習と評価には日本音響学会の音素バランス文からなる研究用連続音声認識データベース (ASJ-PB) , および、新聞読み上げ音声コーパス (ASJ-JNAS) を用い、男性話者・女性話者・混合話者 (それぞれ100名10発話) の不特定話者モデルを用いた連続音素認識を行った。モデルは GMTK⁵⁾ により作成した。テスト音声データのサンプリング周波数は16 kHz である。音響特徴量は0次元を除いた MFCC 12次元と、その Δ と $\Delta\Delta$ の計36次元である。フレームシフトは10 ms、フレーム幅は25msとした。状態には43音素各1状態を用い、BMMの依存関係は各ノード最大5つとし、過去5フレームまで探索を行った。結果は図4に示す。

この実験において相互情報量を用いて構築したBMMから男性で0.6%、女性で0.6%、混合で1.3%向上し、多重解像度独立性検定を導入した本手法の有効性が示された。これは、相互情報量のような分布を仮定した場合には検出できない認識に有効な相関関係を、多重解像度独立性検定を用いてノンパラメトリックに独立性を計算したことで検出することが可能になったからであると考えられる。

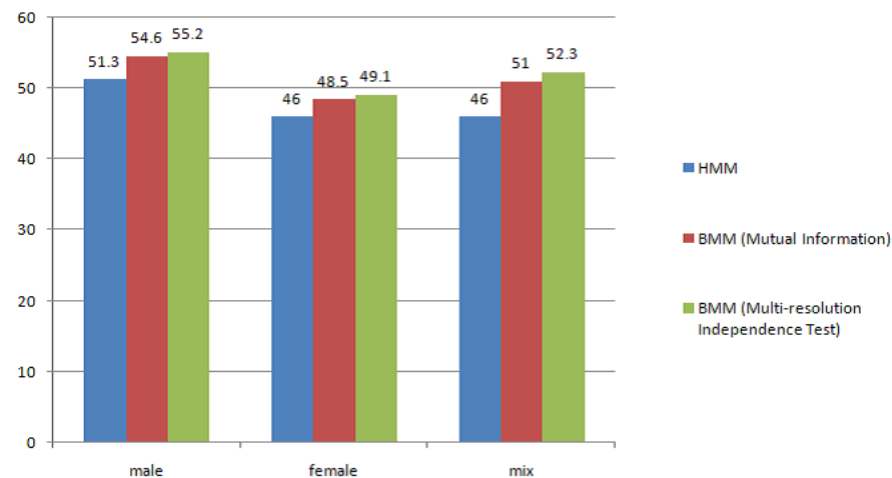


図4 音素認識率 [%]

Fig.4 Phone recognition rate [%]

5. おわりに

BMMの構築法として相互情報量に代わる多重解像度独立性検定を導入した構築法を提案した。GMTKを用いてモデルを実装し、JNASを用いた連続音素認識実験によってその有効性を確認した。今後、ケンドールの順位相関係数、スピアマンの順位相関係数などの他の相関基準を用いた構築法との比較や、Pairwise アルゴリズムに代わる遺伝的アルゴリズムや焼きなまし法などのベイジアンネットワーク構築法の導入を行い、更なる精度の向上を目指す。

参 考 文 献

- 1) J.A. Bilmes, Buried Markov models: a graphical-modeling approach to automatic speech recognition, Computer Speech and Language, Volume 17, Issues 2-3, pp. 213-231, 2003.
- 2) D. Margaritis *et al.*, A Bayesian Multiresolution Independence Test for Continuous Variables, Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence(UAI), pp. 346-353, 2001.

- 3) J. L. Myers, A. D. Well, Research Design and Statistical Analysis (second edition ed.), pp. 508, 2003.
- 4) H. Abdi, Kendall rank correlation, Neil Salkind (Ed.), Encyclopedia of Measurement and Statistics, 2007.
- 5) J.A. Bilmes and G. Zweig, The Graphical Models Toolkit: An open source software system for speech and time-series processing, ICASSP, Orland, 3916-3919, 2002.