2F4-OS-01a-7

多層マルチモーダルLDAと強化学習による意味理解に基づく行動決定

Action Decision Based on Understanding Using Multilayered Multimodal LDA and Reinforcement Learning

長井 隆行[*1] Takayuki Nagai
中村 友昭[*1] Tomoaki Nakamura
アッタミミ ムハンマド[*1] Muhammad Attamimi
持橋 大地[*2] Daichi Mochihashi
小林 一郎[*3] Ichiro Kobayashi
麻生 英樹[*4] Hideki Asoh

[*1]電気通信大学 The University of Electro-Communications
[*2]統計数理研究所 The Institute of Statistical Mathematics
[*3]お茶の水女子大学 Ochanomizu University
[*4]産業技術総合研究所 National Institute of Advanced Industrial Science and Technology

Multilayered multimodal latent Dirichlet allocation (mMLDA) is an extended version of the original LDA. Since the mMLDA has multi-layers, it can probabilistically represent various kinds of concepts and relationship among them. Moreover language can be represented by the model in the same framework. However, the model is just a representation of knowledge and the usage of the model for selecting actions is an open problem. It is a very interesting problem as the model might help to reveal how the acquired concepts are used for action planning and decision making. In this paper, we examine the integration of the mMLDA and reinforcement learning. This is possible because time expansion of the mMLDA can be thought as a partially observable Markov decision process. We discuss the learning and planning methods for the integrated model.

1. はじめに

人間がどのように行動を学習し，計画・決定しているのか，またそうした行動計画・決定と言語理解や思考などの高次機能がどのように結びついているのかは非常に興味深い問題である．ロボットの知能を考える上でも，こうした仕組みを考え実現することが非常に重要なのは明らかであろう．

行動計画や行動 (意思) 決定の問題は，モデル化も含めて従来多くの研究がなされてきた．特に強化学習の枠組みは，試行錯誤から最適な行動を決定する問題を考える上で重要である [Sutton 98, 高橋 00, 田口 05]．近年の脳イメージング研究は，皮質と大脳基底核のループ回路が強化学習の基盤となっていることを明らかにしている [花川 08, 久保田 07]．皮質と大脳基底核のループは並列的に複数の回路が存在し，運動学習から高次のプランニング，言語や社会性に関する学習など，その影響は非常に広範に及んでいる．このことは，前頭前野が階層的に上位の中枢としてカテゴリ化やプランニング，意思決定など高次の情報処理を担っていることを考えれば，自然であるように思われる．このループに，様々なレベルでの学習やプランニング，意思決定，言語などの結びつきを考えるためのヒントがある．

行動決定においては，センサ情報に基づく即時的なものから，記号のような抽象度が高く汎用性の高い仕組みを使った中長期的なプランニングに基づいたものまで様々考えることができる．これは，モデルフリーの行動決定とモデルベースの行動決定と言い換えることもできる．またこれらは，どちらが良いという問題ではなく，我々人間はこれらを共に適切に利用していると思われる．例えば，未知の環境では即時的で反射的な行動決定がベースとなり，良く知っている環境では，学習したモデルを用いた予測やプランニングに基づく行動決定がなされるであろう．さらにこうしたモデルに基づく行動計画能力は，高次の記号操作による思考につながるかもしれない．こうした行動決定のフレームワークは，独立ではなく相互に依存しつつ一つの大きな枠組みの中で実現されていると考えられる．いずれにしても，皮質と大脳基底核のループのような回路を参考にしつつ知能のモデルを検討することで，様々なレベルのプランニングや行動決定，言語などを統一的に扱うことができる枠組みを構築することが本研究の最終的な目的である．

知能のモデルやアーキテクチャは様々なものが提案されているが，運動制御のレベルから抽象的な言語のような記号操作までを統一的に扱っているものはほとんど存在しない．これは，それぞれが知能のある一部分に焦点を当てているためであると言える．我々のグループでは，記号創発ロボティクスの視点から言語を含むロボットによる実世界の理解について検討を続けてきた．これらの取り組みでは，ロボットが経験によって取得するマルチモーダル情報をカテゴリ分類し，概念をボトムアップに形成することを基盤としている [Nakamura 11]．ボトムアップに獲得した概念を用いることで，ロボットは目前の観測データから，モダリティーを超えた未観測情報を予測することが可能となる．我々は，この予測こそが理解であると考えており，言語も同じ枠組みで理解したり生成したりすることができると考えている．ただしこのモデルは，知識の確率的表現であり，行動を計画したり決定する仕組みは含んでいない．そもそもロボットはどのように経験し，どのようにデータが収集されるのか，逆に獲得した知識が行動決定にどのように利用されるのか？これを考えることで，試行錯誤による運動学習から，概念・言語獲得，言語・実世界の理解，行動計画などが一つの枠組みで結びつくと考えている．

そこで本稿では，多層マルチモーダル LDA (mMLDA) と強化学習を統合した枠組みを提案し，概念学習，知識獲得と様々なレベルでの行動決定について検討する．mMLDA はマルチモーダル LDA の拡張であり，複数の概念とその概念間の関係

連絡先: 長井隆行，電気通信大学 情報理工学研究科 知能機械工学専攻，〒 182-8585 東京都調布市調布ヶ丘 1-5-1，tnagai@ee.uec.ac.jp

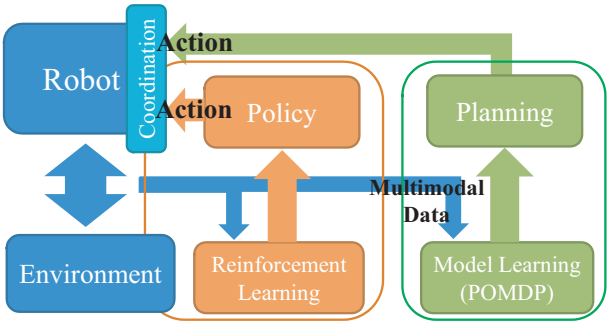

図 1: 提案するモデルの全体像

を確率的に表現することができる [アッタミミ 14]．またこの枠組みで，言語も同様に扱うことができる [Attamimi 14]．しかし，mMLDA で表現されている概念を利用した理解に基づく行動決定については，これまで議論されていない．本稿では，mMLDA を時間的に接続したものが部分観測マルコフ決定過程 (POMDP) であるという視点から，強化学習を用いた行動決定を検討する．ここでのポイントは，次の 3 つである．

一つ目は，強化学習によって得られるマルチモーダルデータの時系列を使って，どのように概念を形成するかという問題である．本稿では，これが時空間的な分節化とカテゴリ分類の問題であると捉え，mMLDA を時間発展させた POMDP によってモデル化することでこの問題を解決することを提案する．

二つ目は，モデルベースの行動決定と，モデルフリーの行動決定をどのように統合するかという問題である．ロボットが未知の環境で行動する場合には，試行錯誤する必要があり，既知の環境ではモデルを使って行動を計画することができる．これを実現するためには，現在の環境が未知であるのか既知であるかのメタ認識が必要である．この点については，文献 [星野 11] において，オープンエンドな知能を実現するためにこうした仕組みが必要であり，これをどのように実現するかが議論されている．本稿では，概念形成を考えているため，より問題が複雑となる．概念は汎化された状況のカテゴリであり，これにより未知の状況に対しても対応できる可能性がある．従って，未知と既知の判断は確率的かつ階層的に行われるべきであり，その点を考慮した仕組みが必要である．また学習はオンラインで進むため，試行錯誤中に何かに気づき，プランニングに移行するなどダイナミックなプロセスでもある．本稿では最初のステップとして，強化学習による行動選択によって収集した情報を用いて概念を形成し，その時空間的に分節化・カテゴリ分類された概念を用いて行動計画を実現することを検討する．その先の複雑な統合問題は，今後の課題としたい．

三つ目は，言語理解や操作と行動計画・決定とのつながりに関するものである．mMLDA は，文法を含む言語の獲得を可能とするモデルであるため，本稿で提案する仕組みを拡張することで，例えば自分の行動を言語化したり，言語による命令を実行することが自然に実現できると考えられる．また，POMDP に基づく対話の学習の様な枠組みが従来から提案されており，提案する枠組みではそうした対話戦略のようなものを獲得する枠組みを内包していると考えられる．

2. mMLDA と強化学習の統合

2.1 問題設定

前章で述べた問題を解決する一つのアイディアとして，強化学習とモデルベースの学習・行動計画を統合するような枠組

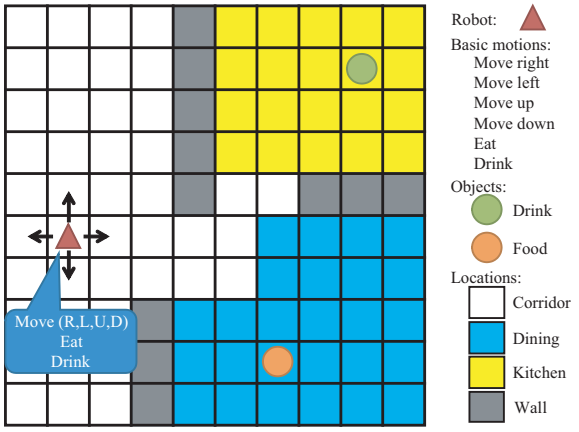

図 2: 具体的な問題の例

みを考える．本稿で提案するモデルの全体像を，図 1 に示す．この図におけるモデルベースの学習には，mMLDA を時間的に接続した POMDP を用いることを想定している．まず，本稿で扱う問題を明確にするために，具体的な問題設定を行う．ここでは，ロボットが図 2 のグリッドワールドの中で，自身の内的欲求を満たすように行動することを考える．ロボットには基本的な動作として，上下左右への移動と，食べる動作，飲む動作の 6 つが備わっている．グリッド内はどこでも移動できるが，壁のグリッドへは移動できない．また，グリッドの各色は，想定した場所を意味している．例えば，黄色はキッチンであり，その中のどこかに飲み物が置かれている．ロボットは，内的状態が「のどの渇き」であった時，試行錯誤することで最終的に飲み物のあるグリッドで飲む動作をすることができれば報酬が与えられる．内的状態が「空腹」であれば，最終的に食べ物のあるグリッドへ移動し，食べる行動をすることで報酬が与えられる．壁には移動できないため，移動しようとすると負の報酬が与えられる．この際，各グリッドを状態とした試行錯誤による学習は，一般的な強化学習である．

本稿で提案するのは，この強化学習の際に得られるデータを mMLDA でカテゴリ分類し，さらに時間的に接続することで，高次の概念を獲得し，それを使って状況理解に基づく行動計画・決定を実現することである．これは，強化学習によって学習される状態価値だけでは実現できない．ここで「理解」とは，概念を通した予測であり，例えば，黄色い領域には飲み物がある可能性が高いといったことが予測できることを意味している．また mMLDA は言語との結びつきを学習することもできる枠組みであるため，黄色い領域を「キッチン」と呼ぶことを学習すれば，キッチンという記号を使った高次の推論が可能となる．さらには，自身の行動を文章として発話したり，自然言語による発話や命令を理解して行動することができる可能性をもった枠組みとなっている．

ここで述べた具体的な問題はあくまで例であり，問題の本質はそのフレームワークの実現にあることは言うまでもない．

2.2 多層マルチモーダル LDA

多層マルチモーダル LDA (mMLDA) は，下位層に物体，動き，場所などの下位概念を表現するマルチモーダル LDA (MLDA) を，上位層にそれらを統合する MLDA を配置した階層的な構造をもつ確率モデルである．これにより，動き，場所，物体など各々のカテゴリ分類を行うと同時に，それらの概念間の関係を教師なしで学習することができる [アッタミミ 14, Attamimi 14]．図 3 に，mMLDA のグラフィカルモデルを示

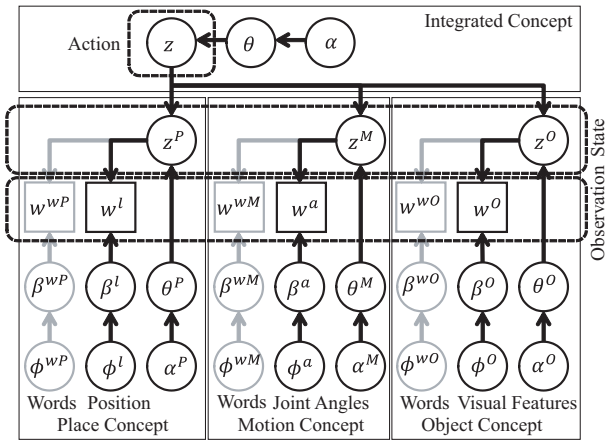

図 3: mMLDA のグラフィカルモデル

す．図 3 において，z は統合概念を表すカテゴリであり，z^O，z^M，z^P はそれぞれ下位概念に相当する，物体，動き，場所カテゴリである．上位カテゴリ z は，下位カテゴリ間の関係性を捉えており，ロボットの行動を表現することになる．w^o，w^a，w^x は観測データであり，それぞれ，物体情報，ロボットの動き，位置情報である．

2.2.1 下位概念

物体情報として，物体番号ヒストグラム $w^o = \{o_1, o_2, \cdots, o_{N_o}, o_{N_{o+1}}\}$ を用いる．ただし，N_o は物体数を表している．o_* は 0 または 1 の値をとり，物体番号 k の物体が観測された場合 o_k が 1 となり，物体が観測されていない場合 N_{o+1} が 1 となる．動き情報としても同様に，基本的な動きに付与されたインデックスのヒストグラムを用いる．場所情報としては，グリッドの位置と代表位置との距離をヒストグラムの形で表現した $w^l = \{l_1, l_2, \cdots, l_{N_l}\}$ を用いる．ここで，N_l は代表場所数を表している．これらは上述の問題設定に従ったものであり，実際にはセンシングに基づく特徴ベクトルなどを用いることも可能である．

2.2.2 統合概念とパラメータ推定

mMLDA では，各概念を表す隠れ変数 z，$z^C \in \{z^O, z^M, z^P\}$ を同時に学習する．学習にはギブスサンプリングを用い，各概念を表すカテゴリ z，z^C を，観測データ $w^m \in \{w^o, w^{wO}, w^a, w^{wM}, w^l, w^{wP}\}$ を用いてサンプリングする．サンプリングには，$\theta, \theta^C, \beta^m$ を周辺化した事後分布を用いる．

さらに，学習モデルを用いることで，物体や動きの認識だけでなく，概念間の予測も可能となる．

2.3 強化学習

本稿では，強化学習として Q 学習を想定しているが，他の学習手法を利用することも可能である．基本的な動作の学習メカニズムとして，MOSAIC 強化学習 [鮫島 01] などを利用することも考えられる．

2.4 mMLDA の時間発展

図 3 を見ると，ロボットの行動を表現する上位層と，下位の各概念の組み合わせで表現される隠れ状態，及び観測データによって mMLDA が構成されていることが分かる．従って，これを時間軸方向に接続することで，POMDP と等価なモデルを構築することができる．図 4 に，この様子を示す．図の左側は，mMLDA の事前分布やハイパーパラメータを省略し，行動，状態，観測で表現したものであり，右側はこれを時間的に

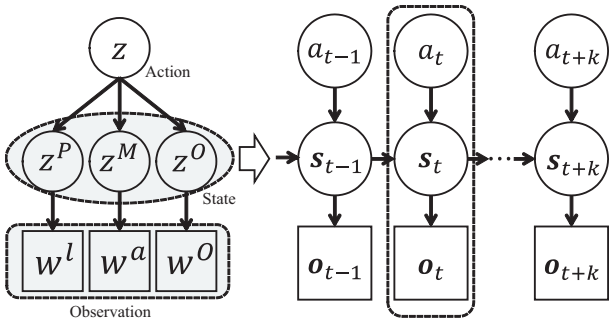

図 4: mMLDA の時間発展 (事前分布やハイパーパラメータは簡単のため省略している)

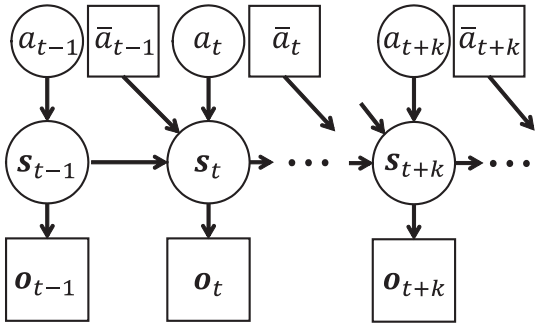

図 5: POMDP のグラフィカルモデル

接続したものである．これは，Input/Output HMM と見なすこともできるが，この段階で行動 a_t は隠れ変数であり，観測できないことに注意が必要である．つまり，ある状態においてどのような行動を取ったのかは確率的な予測であり，観測 $\boldsymbol{o}_t$ から推論する必要がある．

状態 $\boldsymbol{s}_t$ も同様に隠れ変数であり，各下位概念の組み合わせになっている．このように状態を下位概念の全ての組み合わせと考えることで，このモデルの学習は一般的な HMM の学習 (EM アルゴリズム) として定式化できる．一方，ノンパラメトリックベイズモデルである階層ディリクレ過程-HMM (HDP-HMM) [Beal 01] を用いることで，状態数をデータから推定し，自動的に状態空間を構築することも可能である．

ここで更にロボットによる行動決定を考慮すると，図 5 のような POMDP を描くことができる．この図において，a_t は時刻 t にロボットが取り得る行動の確率的な予測を表しているのに対し，$\bar{a}_t$ は実際にロボットが取る行動を表している．ロボットは行動計画によって行動 $\bar{a}_t$ を決定し，$\bar{a}_t$ は次の時刻 $t+1$ の状態に影響を与えることになる．

3. 学習と行動計画・決定

3.1 学習 (概念獲得)

概念学習の問題は，強化学習と同時に行われる時空間分節化・カテゴリ分類の問題であると言える．本稿では簡単のために，強化学習によって蓄積されたマルチモーダルデータを使ってバッチ学習することを考えるが，本来は強化学習と並列かつオンラインでモデルの学習を行い，その時点でのモデルによる行動計画と強化学習の政策を協調させて行動決定すべきであろう．これについては今後の課題とし，ここでは POMDP の学習について考える．

まず mMLDA の学習は，データ全体を mMLDA によって

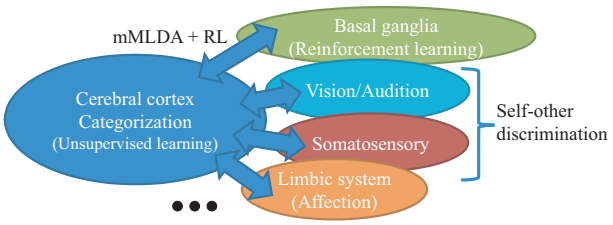

図 6: カテゴリ分類と様々な情報の統合

階層的にカテゴリ分類することで実現される．これが空間的な分節化に相当している．この後に，HDP-HMM (もしくは HMM) の学習を行う．mMLDA によってある種記号化された時系列データを，時間的に分節化しつつカテゴリ分類するのが，HDP-HMM のパラメータ推定であると言える．

3.2 行動計画

ここでは，図 5 のモデルから，どのように行動を計画するかについて考える．入力される信号は，現在時刻 $t = 0$ の観測 $\boldsymbol{o}_0$ であり，最終的な目的状態へ遷移するための最短の行動系列 $\boldsymbol{a} = \{a_0, a_1, \cdots, a_n\}$ を求めることが目標となる．この問題は，時間 t，状態 $\boldsymbol{s}_t$，行動 $\bar{a}_t$ を軸としたトレリス空間で，最尤となるビタビパスを求める問題と考えることができる．ただし，観測データは実際に行動を取らない限り得られないため，現在の観測データのみを考慮する．トレリス上で目標となる状態にたどりついたとしても，その確率が同じ時間における他の状態に存在する確率に比べて小さい場合には，目標の状態にたどりついていない可能性が高い．そこで，目標の状態に他の状態に比べて十分に高い確率で存在する場合に，行動計画が終了したとしてビタビパスをバックトラックすることとする．

以上の手法によって，行動系列 $\boldsymbol{a}$ を計画することができる．しかし，実際には行動 a_t は概念としての行動であり，例えばロボットが「廊下を移動する」といったものである．実際にロボットが行動するためには，具体的な動作を決定する必要がある．つまり，移動という上位の行動概念ではなく，上下左右のどの方向に移動するかという具体的な動作を推定しなければならない．

3.3 動作の決定

決定すべき具体的な動作は，図 3 における w^a に相当する．これは本来可観測であるが，将来の行動計画においては未観測であり，観測情報より推定する必要がある．つまり，$\bar{w}_t^a = \mathrm{argmax}_{w^a}\, p(w^a | z_t, z_t^P, z_t^M, z_t^O)$ を解けばよく，これは mMLDA の枠組みで計算可能である．また，他の観測データについても同様に予測することができるため，実際に行動を行った後に，予測される観測データ・状態と実際の観測データ・状態とのずれを計算することができる．このずれが大きい場合にはリプランニングを行い，新たな計画に従って動作を決定する．

4. 議論

本稿で提案したモデルでは，試行錯誤に基づく強化学習から階層的に概念を形成し，概念を通した予測に基づく行動計画・決定を行うことができる．我々の「理解」に対する定義は，ボトムアップに形成した概念に基づく予測であり，その意味において，理解に基づく行動決定が実現できると言える．紙面の都合上，シミュレーション結果は割愛するが，前章で述べた問題設定のシミュレーションによって，キッチンやダイニング，廊下といった場所概念や，移動，食べる，飲むといった動作概念が形成され，上位概念において，「廊下を移動する」や「キッチンで飲む」といった行動概念が形成されることを確認した．また，こうして形成された概念を基盤としたモデル (POMDP) を用いて，内部欲求に基づき行動を計画し，例えば実際に飲む行動 (キッチンに移動して飲み物を飲む) を計画・実行することも可能である．

今後は，提案したモデルを定量的に評価しつつ，実際のロボットへ搭載することを検討したい．また，本稿では簡単化のために考えなかった問題を検討する必要がある．重要なのは，並列に学習した結果をオンラインでいかに協調させるかであり，活用と探索のトレードオフの問題もこれに関連する．

言語を考えることも今後の課題ではあるが，mMLDA では言語を扱うための検討がすでに進んでおり [Attamimi 14]，その枠組みをそのまま利用することができると考えている．つまり，言語理解に基づく行動や，自身の行動の言語化，言語的思考による行動計画・決定などが可能である．

さらには，感情や他者との関わりなども検討したいと考えている．図 6 に示すように，様々な種類の情報を皮質で教師なし学習し，この結果が利用されるような構造を考えており，これは mMLDA を基盤として実現することが可能である．更なる階層化 (深層化) も興味深い今後の課題である．単純なタスクではなく，より現実的で複雑なタスクを考えることで，モデルの適用範囲がどこまでかを明らかにする必要がある．

謝辞

本研究は，JSPS 科研費 26280096 の助成を受けて実施したものである．

参考文献

[Sutton 98] R.S. Sutton, A.G.Barto, Reinforcement Learning, MIT Press, 1998 （三上ほか 訳：強化学習，森北出版, 2000）

[高橋 00] 高橋, 浅田, “複数の学習器の階層的構築による行動獲得”, 日本ロボット学会誌, vol.18, no.7, pp.1040-1046, 2000

[田口 05] 田口, 桂田, 新田, “並列学習を利用した対話戦略の獲得”, 人工知能学会全国大会, 3E1-04, 2005

[花川 08] 花川, “行動制御における大脳基底核-皮質系の役割：脳機能イメージングからの知見”, ロボティクス・メカトロニクス講演会, pp.1-4, 2008

[久保田 07] 久保田, 酒田, 松村 編, 学習と脳, サイエンス社, 2007

[Nakamura 11] T. Nakamura, T. Araki, T. Nagai, N. Iwahashi, “Grounding of Word Meanings in LDA-Based Multimodal Concepts,” Advanced Robotics, 25, pp.2189-2206, 2011

[アッタミミ 14] アッタミミ, ムハンマド, 阿部, 中村, 船越, 長井, “多層マルチモーダル LDA を用いた人の動きと物体の統合概念の形成”, 日本ロボット学会誌, vol.32, no.8, pp.89-100, 2014

[Attamimi 14] M. Attamimi, M. Fadlil, K. Abe, T. Nakamura, K. Funakoshi, T. Nagai, “Integration of Various Concepts and Grounding of Word Meanings Using Multi-layered Multimodal LDA for Sentence Generation,” in Proc. of IROS, pp.3005-3011, 2014

[星野 11] 星野, 河本, 野田, 佐部, “自己調整学習メカニズム：オープンエンドな環境で発達するエージェントの自律学習行動原理”, 日本ロボット学会誌, Vol. 29, No. 1, pp. 77-88, 2011

[鮫島 01] 鮫島, 銅谷, 川人, “強化学習 MOSAIC: 予測性によるシンボル化と見まね学習”, 日本ロボット学会誌, vol.19, no.5, pp.551-556, 2001

[Beal 01] M.J. Beal, Z. Ghahramani, C.E. Rasmussen, “The infinite hidden markov model”, Advances in neural information processing systems, pp. 577–584, 2001