LJ-011

ベイジアンネットワークを用いた感性会話ロボットのための対話者感情の推定法

A Method of Inferring Dialogist's Emotion for Sensitivity Robots using Bayesian Network

趙 章植 † 加藤 昇平 † 加納 政芳 ‡ 伊藤 英則 †
Jangsik Cho Shohei Kato Masayoshi Kanoh Hidenori Itoh

1 はじめに

ペットロボットや会話ロボットなど多くのエンタテインメントロボットが開発されており，人間とのコミュニケーションを目的としたロボットの研究が盛んに行われている．ロボットが人間とより豊かにコミュニケーションするためには，単に相手の発話音声やその意図を理解するだけでなく，お互いの感情や情動を把握し，これらを有効に利用して感情豊かに振る舞うことが要求される．このように，エンタテインメントロボットには，「相手の感情を推定する」「自己の感情を生起させる」「相手に感情を表出する」ための３つの感情制御機能が必要となる．本論文では，「相手の感情を推定する」機能を実現するための一手法を提案する．

人間は，会話中の相手が発する表情，すなわち，発話音声の様々な特徴や顔表情，あるいは，身ぶりなどから会話相手の感情や情動を推定している．ここでは，話者の発話音声に着目する．発話音声からの感情推定については，音声に含まれる韻律的特徴が感情表現に関する多くの情報を持っていることがこれまでの研究で明らかになっている [1, 2, 3, 4]．重永らは [2]，音声韻律的特徴量の平静からの「ずれ」を使う正規化法により音声韻律的特徴量の各感情判別への寄与の程度を明らかにしている．白澤ら [4] は多変量解析の手法を用いて音声韻律的特徴量を解析し特徴量の主成分のマハラノビス距離を計算することで高い精度で各感情を分類することに成功している．また森山ら [3] は，ファジー制御を用いることで音声に含まれる情緒性を評価するシステムを提案している．しかしながら，ロボットと人間が会話する実環境での実時間情報処理を考えると，環境音などから受けるノイズの影響やロボット内部のモーター音等の干渉などが想定されるため音声解析の理想状態からは程遠く，一部の韻律特徴量が欠損したり，抽出された数値の信頼性が低下するなど，感情推定のための情報処理には不確実性が大きく伴う．

一方，ベイジアンネットワークとよばれる確率的な知識表現と推論の枠組が人工知能の分野で研究されてきており，不確実な知識の下での知識表現能力と推論能力の高さから，近年，故障診断，対話プランのユーザモデリング [5]，自然言語処理 [6] など，様々な分野へ応用されるようになってきた．

そこで本論文では，感性ロボットの対話者感情推定手法として，音声データに含まれる韻律特徴に着目し，ベイジアンネットワークを用いて感情推定のための知識と推論モデルをロボットに持たせる方法を提案する．

2 感性会話ロボット ifbot

我々は，音声会話処理と表情表出の基本機能を持った感性会話ロボット ifbot を開発し [7, 8, 9]，ifbot を用いて感情制御に関する研究を行っている．図 1 に感性会話ロボット ifbot[10] の外観を示す．ifbot は二本の腕を持ち，足の代わりに車輪により移動する全長 45cm，体重 7kg のロボットである．同ロボットには音声合成および音声認識の機能が搭載されており，人間と会話することができる．また，感情を伴う対話を目的とした表情表出機能として 10 のモータおよび合計 104 の LED を持つ．モータは首に２つ，目および瞼にそれぞれ左右独立して２つづつ搭載している．LED は頭部，口部，目部，頬部，涙部及び耳部の各部位に配置しており，各部位の LED が，幾つかのパターンで発色することにより「喜び，怒り，悲しみ，驚き」など 40 種類以上の表情を表すことができる．また，最大 10 名の対話者を認識するためのカメラ及び画像処理の機能を持つ．

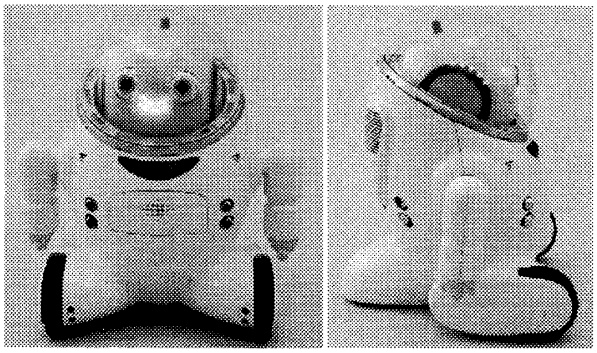


図 1 感性会話ロボット ifbot

本研究の目的は ifbot に人間との感性豊かなコミュニケーション能力を持たせることにある．そのためには，上記の機能に加えて人間の感情を把握するメカニズムが必要となる．そこで本論文では，対話者が発話した音声から対話者感情を推定するための感情推定知識をベイジアンネットワークとしてモデル化する．そして，ifbot の会話処理のプロセスにベイジアンネットワークモデルを用いた対話者感情推論機能を組み込むことを提案する．図 2 に本研究で提案する対話者感情推定手続きの概要を示す．本論文では，発話音声の韻律特徴解析に基づいた感情推定のためのベイジアンネットワークモデルの構築方法について述べる．

3 ベイジアンネットワーク

ベイジアンネットワーク（BN）は，複数の確率変数の間の定性的な依存関係を非循環有向グラフ（DAG）により表現し，個々の変数の間の定量的な関係を条件付確率で表した確率モデルである [11]．確率変数をノードとし，変数間に確率的依存関係が強いと判断される場合に対応するノード間に有向リンクを付ける．依存関係を確率的相関と同一視した場合，N 個の確率変数 $(X_1, ..., X_n)$ の同時確率分布 P は次式で表現される．

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|Pa(X_i)). \quad (1)$$

ここで，$Pa(X_i)$ は確率変数 X_i の親ノードを表す．式 (1) は，各ノード X_i が $Pa(X_i)$ にのみ依存し，X_i から辿って到達できるノードを除いた他のノードとは条件付独立となることを表している．

親ノードがある状態 $Pa(X_i) = x$（x は親ノード群の各値で構成したベクトル）の下での n 通りの離散状態 $(y_i, ..., y_n)$ を持つ変数 X_i の条件付確率分布は $p(X_i = y_1|x), ..., p(X_i = y_n|x)$ となる．これを各行として，親ノードがとりうる全ての状態

† 名古屋工業大学, Nagoya Institute of Technology
‡ 中京大学, Chukyo University

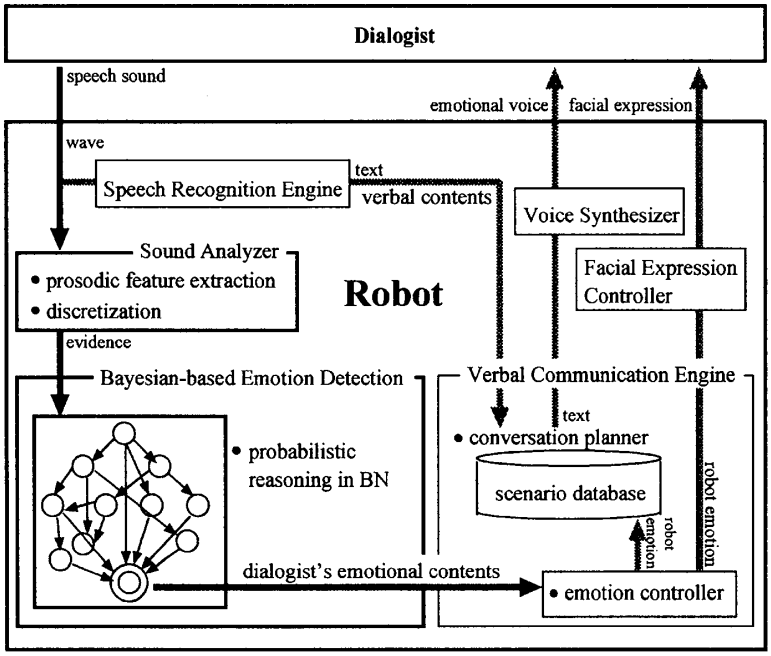


図 2　ロボットの会話処理部におけるベイジアンネットワークを用いた対話者感情の推定手法

$Pa(X_i) = x_1, ..., x_m$ のそれぞれについて列を構成した表の各項目に確率値を定めたものが X_i についての条件付確率表（CPT）である．これにより，確率変数間の確率的な依存関係がモデル化できる．ベイジアンネットワークを用いて知識をモデル化することで，知識の記述量及び計算量が大幅に削減される．また部分的な証拠からでも確率的に推論できる長所を持つ．このため本研究では，ロボットに搭載する感情推定のための知識モデルとして効率とロバスト性を得ることが可能なベイジアンネットワークを応用する．

4　音声の韻律特徴を用いた感情推論器の学習

本研究では，対話者が発話した音声から対話者の感情を推定するためにロボットに与えるべく音声に対する感情推定知識をベイジアンネットワークとしてモデル化した．本節では，モデル構築の流れについて概説する．

4.1　音声資料

使用する音声資料は，感情表現がなされている必要がある．本研究では TV ドラマ，映画から俳優が感情を込めて発話したフレーズを抽出し，「怒り」「悲しみ」「嫌悪」「恐怖」「驚き」「喜び」の 6 種類（以降，6 感情）に分類した．それらの中から，聴取実験により感情が適切に表現されていると判断された音声資料をサンプルデータとする．

4.2　特徴量の抽出

音声は，3 つの要素（韻律，音質，音韻）から成り立っている．この中で，韻律的特徴が人間の感情表現に最も関連することが過去の様々な研究から明らかになっている（例えば文献 [1, 2, 3, 4]）．そこで本研究では，音声資料からピッチ構造を反映する「基本周波数」（$F0$），振幅構造を反映する「短時間パワー」（PW），及び時間構造を反映する「1 モーラあたりの発話継続時間」（Tm）をそれぞれ計測する．図 3 に，ある女優が「わかります」と発話した音声資料に対する特徴量の抽出例を示す．$F0$ 及び PW に関しては，平均，最大，最小，標準偏差を抽出する．このとき，短時間分析におけるフレーム長を 23ms（250 samples），フレーム周期を 11ms とし，窓関数として Hamming 窓を使用した．以上により求めた 9 個の特徴量に加えて，発話者の性別（SE）を加えた 10 個の特徴量をベイジアンネットワークの確率変数とする．

4.3　音声特徴の量子化

4.2 節で述べた音声資料から抽出した各特徴量の統計量は連続値を取るため，離散状態を扱うベイジアンネットワークに適用するためには特徴量の量子化が必要である．量子化数やその閾値はベイジアンネットワークの学習に必要なデータ量や学習データにおける各特徴量の存在分布により適切に設定する必要がある．

4.4　モデルの構造決定

本研究では，学習データに含まれる目標属性（6 感情）と属性（音声韻律特徴）との間の依存関係を表現するために属性間の結合とその強さ（CPT）を学習することでベイジアンネットワークの構造を決定する．学習方法として，本研究では情報理論的妥当性がありデータへの過度なフィットを回避することで予測精度の高いモデルが学習可能な BIC（Bayesian Information Criterion）に基づくモデル選択手法を採用する．M をモデルとし，$\hat{\theta}_M$ を M を表すパラメータ，d をパラメータ数とすると M の評価値 $BIC(\hat{\theta}_M, d)$ は次のように定義される．

$$BIC(\hat{\theta}_M, d) = \log_{\hat{\theta}_M}^{N} P(D) - \frac{d \log N}{2} \qquad (2)$$

ここで D は学習データ，N は D のデータ数を表す．$\hat{\theta}_M$ は最尤法により求めた．D が部分観測の場合には EM アルゴリズム [12] を用いて推定し CPT を補間する．本研究では，BIC が最大となるモデルを求めこれを対話者の感情推定のための知識としてロボットに与える．BIC を最大にするモデルの探索には K2 アルゴリズム [13] を用いた．K2 アルゴリズムではノード間の親子順序を事前知識として与えることで探索空間を制限することが可能だが，ここでは音声の振幅（PW），ピッチ（$F0$），時間（Tm），及び，性別（SE）の 4 グループに分け，グループ内のノードのオーダリングを固定することにより探索空間を軽減させた（図 4）．性別（SE）ノードは最上位に固定しその他 3 つのグループの順列組合せについて学習を行い，学習データを用いた対話者感情の正答率が最も高いモデルを採用することで準最適な構造を決定する．

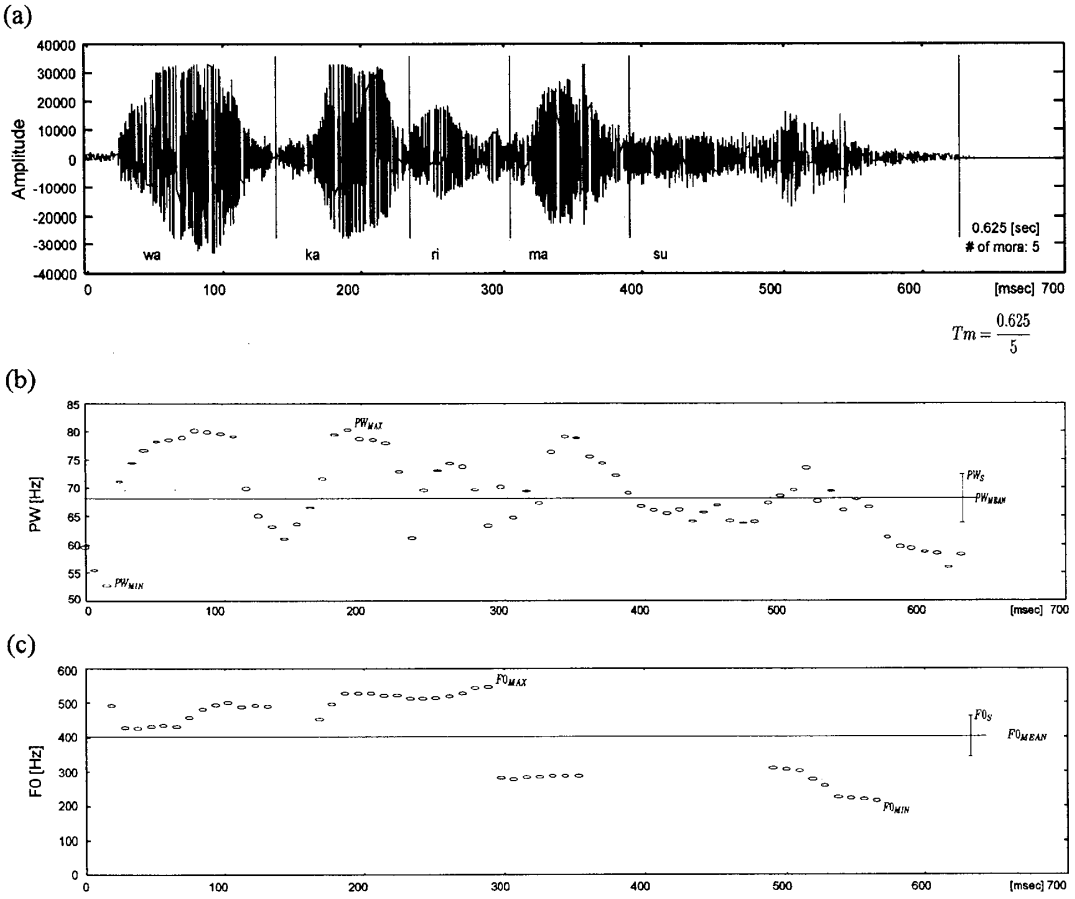
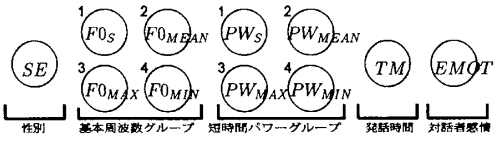


図 3　発話音声「わかります」に対する韻律特徴抽出の例：(a) 音声波形とモーラによる分節結果，(b) 各フレーム毎のエネルギー (*PW*) プロット，(c) 各フレーム毎の基本周波数 (*F0*) の推移



1 $SE \prec F0group \prec PWgroup \prec TM \prec EMOT$
2 $SE \prec F0group \prec TM \prec PWgroup \prec EMOT$
3 $SE \prec PWgroup \prec F0group \prec TM \prec EMOT$
4 $SE \prec PWgroup \prec TM \prec F0group \prec EMOT$
5 $SE \prec TM \prec F0group \prec PWgroup \prec EMOT$
6 $SE \prec TM \prec PWgroup \prec F0group \prec EMOT$

図 4　ノードのグループ化とオーダリング

5　感情推論のアルゴリズム

本提案のヒューマン－ロボットインタラクションにおける実応用を考えた場合，会話相手の性別が認識できなかったり音声特徴の解析が一部失敗したりするなど，不確実性を伴う情報や部分的な観測情報を基に推論を行うことが要求される．ベイジアンネットワークの確率推論は十分な情報がないときでも合理的な意思決定を行うことが可能である．これにより一部の証拠のみが与えられた場合でも確率推論ができる．本研究では，ネットワーク構造に復結合を持つ場合でも効率的に推論を行うことができる Junction Tree[14] を感情推論のアルゴリズムに採用する．

6　感情推定実験

本論文で提案した手法の有効性を確認するため，感情推定実験を行った．まず，4 節で述べたように，6 感情のいずれかにラベル付けされた音声資料を 1600 事例用意し，話者の性別ラベルと音声の韻律特徴を抽出し属性を付与した後，任意に 1400 の学習事例と 200 のテスト事例を作成した．韻律特徴量の量子化数はすべて 5 とした．各属性に対する量子化の閾値は，各量子化レベルに属する学習データの数が均一 (20%) になるように決定した．図 5 に学習事例から作成されたベイジアンネットワークモデルを示す．ここで得られた変数の順序は $F0_S \prec F0_{MEAN} \prec F0_{MAX} \prec F0_{MIN} \prec PW_S \prec PW_{MEAN} \prec PW_{MAX} \prec PW_{MIN} \prec Tm$ であった．

6.1　情報欠損に対する頑健性評価

評価実験はテスト事例から 10 属性すべての証拠が与えられた場合と，6 証拠 ($SE, F0_S, F0_{MAX}, PW_S, PW_{MEAN}, Tm$) のみ，ならびに，4 証拠 ($F0_{MAX}, PW_S, PW_{MEAN}, Tm$) のみが与えられた場合の 3 つについて行った．表 1（左）に感情推定の正答率を示す．全証拠を用いた実験では 6 感情すべてにおいて 55% 以上の正答率で認識された．特に「怒り」の感情について高い認識性能が確認された．なお，「恐怖」と「驚き」の音声データを十分収集できなかったため，これらの感情の正答率が低下した．なお，今回の実験では，様々な俳優から音声資料を抽出したため，俳優による音声特徴の格差が正答率に影響している．一方で，6 証拠および 4 証拠のみを用いた実験では「悲しみ」「嫌悪」の正答率が大きく低下したものの，6 感情を無作為に回答した

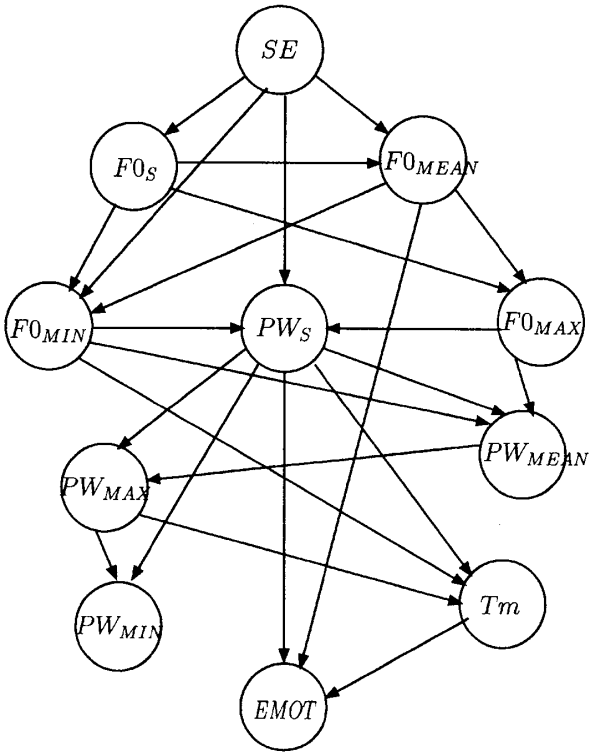


図 5　生成されたベイジアンネットワークモデル

表 1　感情推定実験結果

感情	正答率（%）			
	提案（BN）手法			PCA
	全（10）証拠	6 証拠	4 証拠	
怒り	72.1	54.1	58.1	63.6
悲しみ	64.9	24.3	27.0	27.0
嫌悪	64.1	38.5	35.9	61.5
恐怖	55.6	37.0	37.0	48.1
驚き	59.3	48.1	44.4	44.4
喜び	63.0	40.7	44.4	22.2

場合（16.7%）を下回ることはなく，6 感情すべてが認識されたといえる．このことから提案手法の頑健性が確認された．

6.2　多変量解析手法との比較

多変量解析手法との比較実験を行った．学習データの解析には主成分分析（PCA）を用いて累積寄与率が 9 割超となる第 6 主成分までを選択した．テストデータの感情分類にはマハラノビス距離を用いた判別分析法を採用し，判別方法には 2 次多項式による非線形判別法を用いた．表 1（右）に結果を示す．6 感情すべてにおいて，PCA より本手法の正答率が高いことがわかる．実験で用いた音声資料に関しては「嫌悪」をのぞく 5 感情について大幅な正答率の改善が確認された．このことから提案手法の有効性が確認された．

7　おわりに

本論文では，感性会話ロボットのための対話者感情の推定法を提案した．提案手法は，発話音声の韻律特徴を用いて会話中の相手の感情を推定するための知識と推論をベイジアンネットワークを用いてモデリングする方法である．本手法により，音声情報から対話者の感情を推定することが可能となった．

人間の実環境において自ら移動し働きかけるような感性会話ロボットにおいて，高い精度や頑健性を保った音声解析を実現することは困難であると考えられる．本提案はこのような場合において有効な手法である．今後の課題として，発話者毎の発話特徴の差異を考慮した学習により感情推定を改善したい．そして，言語や表情などを含めた総合的な感情推論器を構築し ifbot への実装を行う．

謝辞

ifbot は株式会社ビジネスデザイン研究所の製品企画・総合プロデュースのもと，特に，形状デザイン，表情制御メカニズムはブラザー工業株式会社，および，表情制御ソフトはブラザー工業株式会社，ロボス株式会社，名古屋工業大学が共同開発した．関連各位に感謝する．本研究の一部は，文部科学省科学研究費補助金基盤研究（C）（課題番号 17500143），および，堀情報科学振興財団によった．

参考文献

[1] K. R. Scherer, T. Johnstone and G. Klasmeyer: “Vocal expression of emotion”, R. J. Davidson, H. Goldsmith, K. R. Scherer eds., Handbook of the Affective Sciences (pp. 433–456), Oxford University Press (2003).

[2] 重永：“感情の判別分析からみた感情音声の特性”, 電子情報通信学会論文誌, **J83-A**, 6, pp. 726–735 (2000).

[3] 森山, 小沢：“ファジー制御を用いた音声における情緒性評価法”, 電子情報通信学会論文誌, **J82-D-II**, 10, pp. 1710–1720 (1999).

[4] 白澤, 山村, 田中, 大西：“音声に込められた感情の判別”, 電子情報通信学会技術研究報告, 第 HIP96-38 巻, pp. 79–84 (1997).

[5] T. Akiba and H. Tanaka: “A Bayesian approach for user modelling in dialog systems”, 15th International Conference of Computational Linguistics, pp. 1212–1218 (1994).

[6] 乾, 徳永, 田中：“確率的制約に基づく発話プランニング”, 情報処理学会研究報告，自然言語処理研究会報告, pp. 25–32 (1994).

[7] 加納, 後藤, 加藤, 中村, 伊藤：“ロボットの混合感情表出のための表情制御手法”, 日本知能情報ファジィ学会誌, **17**, 2, pp. 250–255 (2005).

[8] S. Kato, S. Ohshiro, H. Itoh and K. Kimura: “Development of a communication robot ifbot”, IEEE International Conference on Robotics and Automation (ICRA2004), pp. 697–702 (2004).

[9] 竹内, 酒井, 加藤, 伊藤：“対話者好感度に基づく感性会話ロボットの感情生成モデル”, 第 11 回ロボティクスシンポジア, pp. 74–79 (2006).

[10] B. D. L. C. Ltd.: “*The Extremely Expressive Communication Robot, Ifbot*”, http://www.business-design.co.jp/en/product/001/index.html.

[11] K. B. Korb and A. E. Nicholson: “Bayesian Artificial Intelligence”, Chapman & Hall/CRC (2003).

[12] A. Dempster, N. Laird and D. Rubin: “Maximum likelihood from incomplete data via the EM algorithm”, Journal of the Royals Statistical Society, **B 39**, pp. 1–38 (1977).

[13] G. F. Cooper and E. Herskovits: “A Bayesian method for the induction of probabilistic networks from data”, Machine Learning, **9**, pp. 309–347 (1992).

[14] F. V. Jensen: “Bayesian Networks and Decision Graphs”, Springer-Verlag (2001).