

G-019

不完全データベースを使用した 学習支援システム向け Causal Network の自動構築法

大西 佳亮 †
Yoshiaki Ohnishi

藤原 祥隆 †
Yoshitaka Fujiwara

岡田 信一郎 †
Shinichirou Okada

吉田 秀樹 †
Hideki Yoshida

1. はじめに

幅広い内容を持つ分野の学習を行うには、得意科目を伸ばすよりも苦手科目を克服する方が学習効果が高い。我々はユーザ適応化を実現するため確率的推論技法である Causal Network [1] を用いユーザの苦手な分野を優先して学習させる学習支援システムを提案してきた [2]。これまで Causal Network の構築は教材作成者の専門知識に依存してきたが、教材作成者にとって負担が大きく、また、実際の学習者の苦手分野の傾向を反映していない可能性がある。そこで教材作成者の負担を減らすために、実際の運用で得られた各問題の解答結果をデータベースとし、その不完全なデータベースから Causal Network の構造を推測し、自動構築する方法の研究を行っている。本稿では、データベースをシミュレーションにより作成し、Causal Network の構造の一部分を推測した結果をまとめた。

2. Causal Network

Causal Network(以下 CN)とは、事象をノードとし、事象間の因果関係を表すエッジでそれらを結合して DAG(非循環有向グラフ)を構成し、ある事象について分かった事実から他の事象の確率的状態がどのようになるかを推論する確率的技法である。

図 1 の構造を使い CN の例を示す。ノード間の関連をエッジで表し、各ノードに条件付確率が設定されている。条件付確率 $P(q211=○ | s21=○)$ は S21 が理解していると分かったとき ($s21=○$) の Q211 が正解になる ($q211=○$) 確率を表している。このとき、「Q211 が正解した」という事実が分かったとする。その事実 $P(q211=○) = 1.0$ を CN に入力すると、CN は与えられた条件付確率を元に、S21 を理解している確率 $P(s21=○)$ 、C21 を理解している確率 $P(c2=○)$ 、S22 を理解している確率 $P(s22=○)$ 等を計算で求めることができる。計算手順は省略するが、この場合の計算結果は $P(s21=○) = 0.77$ 、 $P(c2=○) = 0.5$ 、 $P(s22=○) = 0.6$ となる。

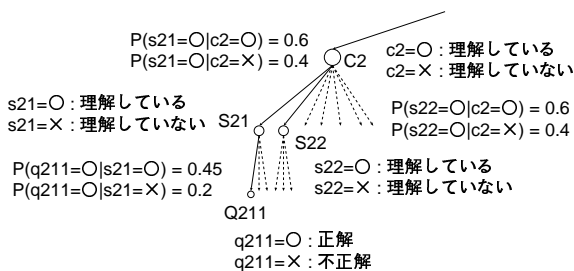


図 1: CN の例

図 2 の構造の学習支援システムでは各ノードを Total, Chapter(C2~C4)、Section(S21~S46)、Question(Q211~Q463) と階層的に分類している。Question は問題を表し 正解/不正解 の 2 つの状態を持ち、それ以外のノードは問題の分野を表し 理解している/理解していない の 2 つの状態を持つ。Total, Chapter, Section の理解している確率を理解度と呼び、理解度からユーザの苦手な分野を推測している。各問題 (Question) は 1 つの分野 (Section) に属しているが、問題によっては複数の分野に関連のあるものもあり、その場合は関連のある分野 (Chapter) と問題 (Question) をエッジでつなぐ。

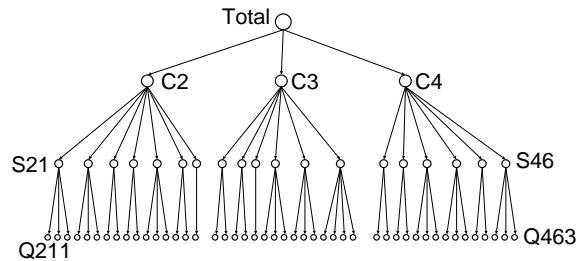


図 2: 題材の構造

3. アルゴリズム

今まで、問題が複数の分野に関連する場合に存在する Chapter - Question 間のエッジを発見するために、教材作成者が全ての問題に対して他の分野との関連を検討していた。しかし、問題数が多く教材作成者の負担が大きいという問題がある。そこで、シミュレーションによって求めた各問題の解答結果をデータベース (以下 DB) とし、その不完全な DB から関連のある Chapter - Question 間のエッジを推測し、CN の構造の一部分を推測する方法を提案する。

3.1 Chapter - Question 間のエッジの推測

新しいエッジの推測には K2 アルゴリズム [3] を改良したものを使用する。K2 アルゴリズムとは、エッジが無くノードだけがあるとき、ノードの状態をデータ値として持つ完全 DB から CN の構造を推測する技法である。改良点は以下の通りである。

- ・ 各ノードに深さレベルを持たせ階層構造にする。
- ・ 基準となる構造 (図 2) が存在し、その構造は変化しない。
- ・ 基準となる構造に含まれていない Chapter と Question 間のエッジだけを推測する。
- ・ 不完全 DB を完全 DB に近似した DB を使用する。

改良した K2 アルゴリズムを使用し、Chapter - Question 間のエッジを発見することができれば教材作成者の負担を減らすことができる。不完全 DB から完全 DB へ

†北見工業大学

の近似方法は以下に示す。

3.2 完全 DB への近似

まず表 1 のような Question にしかデータが無い不完全 DB を用意する。各 Question には教材作成者が設定した 5 段階の難易度があり、難易度が難しい順 (5~1) からその問題の正解率を 0~30%、30~40%、40~60%、60~70%、70~100% とした。例えば、データ数が 100 個あり正解率が 30% ならば、30 個の正解と 70 個の不正解がランダムに並んでいる。

次に、不完全 DB から表 2 のような完全 DB を近似する。全てのノードには難易度に合わせて重み付けをする。今回は難易度 1 から順に 0.8、0.9、1.0、1.1、1.2 とした (Question 以外のノードの難易度は全て 3)。Question 以外のノードは自分に属するノードの数の 1/2 を閾値とし、属するノードの内正解した (又は理解している) ノードの重みを合計し、閾値を越えていればデータ値を理解しているとする。例えば、S21 に属しているノードは Q211、Q212、Q213 の 3 つなので閾値は 1.5 となる。また、Q211 と Q213 が正解していたとすると難易度は各々 4 と 5 なので重みは 1.1 と 1.2 で合計が 2.3 になり閾値を越えているので S21 のデータ値は理解しているとなる。このようにして全てのデータを推測し表 2 のような完全 DB に近似する。

回数	1	2	3	4	5	...
Q211	×	×	×	○	○	...
...			...			
Q463	×	○	○	×	×	...
S21~S46			データ無し			
C2~C4			データ無し			
Total			データ無し			

表 1: 不完全データベース

回数	1	2	3	4	5	...
Q211	×	×	×	○	○	...
...			...			
Q463	×	○	○	×	×	...
S21	×	×	○	○	○	...
...			...			
S46	○	○	○	×	×	...
C2	○	○	×	×	○	...
C3	○	○	×	○	○	...
C4	×	×	×	×	×	...
Total	○	○	×	×	○	...

表 2: 完全データベース

4. 実験

Q232 が属する分野 S23 以外に C4 と関連があると仮定する。このとき、Q232 のデータ値 (正解/不正解) によって変化した S23 の理解度の変化量に応じて完全 DB に近似するときの C4 の閾値を変化させる。このように Q232 と C4 の関連を DB に反映し、改良した K2 アルゴリズムによって Q232 と C4 の間にエッジが推測されれば問題と関連のある分野を推測できたことになる。図 3 に Q232 が C4 と関連があると仮定した場合に推測できたエッジとそのエッジが推測できた確率を示した。

図は基準となる構造のエッジを点線で表し、推測されたエッジを実線で表した。図中に示したエッジは推測できた確率の高いもの 5 つに限定した。

図を見ると、関連があると仮定した Q232 と C4 の間にエッジが推測されたことが分かる。しかし、他にも推測されたエッジが存在し、そのエッジが推測された確率は Q232 のエッジが推測された確率と大きな差は無い。その理由としては、Chapter のデータ値は他の Chapter に属する Question からの影響よりも、その Chapter に属する Section と Question からの影響を強く受けるので、Chapter と Question のエッジを正確に推測するのは困難だと考えられる。それでも、関連があると仮定したノード間にエッジができる確率は相対的に高い確率を示すと予想されるので、ノード間の関連を発見するための参考にする事ができる。

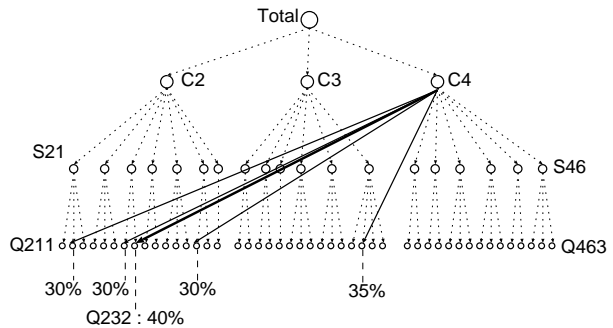


図 3: 推測されたエッジとその確率

5. まとめ

学習支援システムにおいて教材を作成する際に、現在では教材作成者が全ての問題に対してどの分野と関連があるのかを検討しているが、本稿で紹介した方法を用いれば新しくエッジのできた分野と問題との関連を検討するだけで良く、検討する問題数が減るため教材作成者の負担が減ると予想される。

今後は、さらに高い精度で問題と分野の関連を発見するために今回用いた手法を改良し、関連を持つ問題と分野の組を増やした場合も検討する。また、シミュレーションではなく、実際に問題を解いた結果からデータベースを構築し、実際の問題と分野の関連を発見できるかを確かめる必要がある。

参考文献

- [1] Richard E. Neapolitan : “PROBABILISTIC REASONING IN EXPERT SYSTEMS (THEORY AND ALGORITHMS)”, JOHN WILEY & SONS, INC.1990.
- [2] 松澤文太郎, 藤原祥隆, 岡田信一郎 : “ユーザ適応化機能を備えた学習支援システム”, 情報処理学会研究報告, vol.2001, no.122, pp.17-24, Dec. 2001.
- [3] G.F.COOPER and E.HERSKOVITS : “BAYESIAN INDUCTION OF PROBABILISTIC NETWORKS”, Kluwer Academic Publishers(1992).