

ブール多値論理系意味ネットワーク上のクラスタ分割による 知識の局所的一貫性確保

Ensuring Local Consistency of Knowledges by Clustering Semantic Networks on Boolean Multivalued Logic System

情報工学専攻 野崎裕人
Information and System Engineering, NOZAKI Yuto

1 はじめに

ブール多値論理系において、与えられたすべての知識に関して、知識が持つ論理式から計算される実現値が、知識がもつ目標値にできる限り近い値をとれるようにアトム値を求める過程を学習とよび、学習によって求められたアトム値を用いて、任意の論理式の真理値を算出する過程を推論とよぶ [1]。ブール多値論理の次元に対して知識の数が多いときは、間隙は小さくならない。本研究は、述語表現の論理式により記述された知識が与えられたとき、知識を構成する原子論理式および論理式、ならびに知識が持つ目標値をそれぞれオブジェクトまたは矢線に付すことにより、意味ネットワークを構築し、構築された意味ネットワークをクラスタリングによって既定数のクラスターに分割後、各クラスターにおいて間隙の平均を極小化することにより、局所的に一貫した知識を確保し、知識の一貫した範囲内で推論をおこなう方式を提案する。本稿は提案手法および、提案手法によって推論をおこなうシステムの、知識獲得段階からの設計手法について述べ、その設計にもとづいて作成した推論システムによる結果の一例の報告をおこなう。

2 ブール多次元論理

本章では [2] によって定義されるブール多次元論理およびブール多次元論理上の学習と推論について述べる。

2.1 論理値と真理関数

$n \in \{1, 2, \dots\}$ 次元の論理値を、 n 個のビットの並置、すなわち、 $\{1, 0\}^n$ 内の要素として定義する。また、 n 次元の論理値をとりうる変数を、 n 次元の論理変数と定義する。さらに真理関数 T を、 n 次元の論理値に対し実数

$$\text{論理値を構成するビット 1 の個数} / n \quad (1)$$

をとる関数と定義し、この実数を真理値とする。

2.2 論理演算

ブール多次元論理上に定義される論理演算を示す。 $n \in \{1, 2, \dots\}$ 次元の論理値 $p = p_1 p_2 \dots p_n \in \{1, 0\}^n$ 、 $q = q_1 q_2 \dots q_n \in \{1, 0\}^n$ 、に対し以下のような演算を定義する。

否定

論理値 p に対するビットごとの 2 値の否定演算を n 次元の否定演算とし、記号 \neg を用いて $\neg p$ と表記する。ビットごとの 2 値の否定演算は表 1 のとおりである。

論理積

論理値 p および q に対するビットごとの 2 値の論理積演算を n 次元の論理積演算とし、記号 \wedge を用いて $p \wedge q$ と表記する。ビットごとの 2 値の論理積演算は表 2 のとおりである。

論理和

論理値 p および q に対するビットごとの 2 値の論理和演算を n 次元の論理和演算とし、記号 \vee を用いて $p \vee q$ と表記する。ビットごとの 2 値の論理和演算は表 2 のとおりである。

含意

論理値 p および q に対するビットごとの 2 値の含意演算を n 次元の含意演算とし、記号 \rightarrow を用いて $p \rightarrow q$ と表記する。ビットごとの 2 値の含意演算は表 2 のとおりである。

等価

論理値 p および q に対するビットごとの 2 値の等価演算を n 次元の等価演算とし、記号 \leftrightarrow を用いて $p \leftrightarrow q$ と表記する。ビットごとの 2 値の等価演算は表 2 のとおりである。

表 1. 否定演算

p_i	$\neg p_i$
1	0
0	1

表 2. 論理積演算, 論理和演算, 含意演算, 等価演算

p_i	q_i	$p_i \wedge q_i$	$p_i \vee q_i$	$p_i \rightarrow q_i$	$p_i \leftrightarrow q_i$
1	1	1	1	1	1
1	0	0	1	0	0
0	1	0	1	1	0
0	0	0	0	1	1

2.3 論理式

論理演算を論理値または論理変数に適用して合成された結果を論理式とよび, 論理演算を再び論理式に適用して合成された結果も論理式という. 本論文では論理式を構成する論理値および論理変数をアトムとよび, 合成された論理式と区別する.

2.4 述語論理

本研究では, $n \in \{1, 2, \dots\}$ 次元の論理値 p において, ビットが 1 である位置のビットのみを論理値 q から取り出して並置した結果を, $q[p]$ と表記する. ここで p は主語, q は述語とよび, この表記が用いられる論理を述語論理とよぶ.

2.5 知識と間隙

ブール多次元論理では, 知識は論理式 x , 真理関数 T , 目標値 $r \in [0, 1]$ をもちいて $T(x) \approx r$ と表記する. この $T(x) \approx r$ は, 真理値 $T(x)$ が目標値 r にできる限り近い値をとるべきであることをあらわす. また, 各々の知識において, 目標値と実現値の差の絶対値を間隙とよぶ.

2.6 学習と推論

ブール多次元論理における学習とは, 与えられた全ての知識 $T(x) \approx r$ において, 実現値 $T(x)$ が目標値 r にできる限り近い値をとるようにアトムの値を求めることである. また, 本研究では, 知識に関する間隙の全体的な評価を, 間隙の平均をもちいておこなう. よって本研究における学習とは, 与えられた全ての知識に対して, 間隙の平均ができる限り小さくなるようにアトムの値を求めることである.

3 知識獲得および意味ネットワーク構築

本研究ではシナリオベース対話をとおして知識を多数獲得し, それらの知識群を用いて意味ネットワーク

を構築する.

3.1 知識

ブール多次元論理における学習で用いる知識は

$$T(\text{論理式}) \approx \text{目標真理値} \quad (2)$$

と記述できる. 本研究では, 任意のアトムを a, b , 任意の論理式またはアトムを c としたとき, 知識を構成する論理式は述語表現を用いて記述された $(b \rightarrow c)[a]$ と定める. すなわち, 本研究で扱う知識は,

$$T((b \rightarrow c)[a]) \approx \text{目標真理値} \quad (3)$$

である.

3.2 意味ネットワーク

意味ネットワークとは異なる概念間の意味関係をあらわす記述法である. 意味ネットワークは, 概念をノードであらわし, 概念間の意味関係をリンクであらわすことにより, 概念間の意味関係を記述する. ここで, 本研究で用いる意味ネットワークのリンクは, 確信の度合いをあらわす真理値が付加されている. また, 本研究は意味ネットワークの記述を, グラフデータベース neo4j によっておこなう

3.3 シナリオベース対話を用いた知識獲得

シナリオベース対話とは, 対話の要素としてシナリオを定義し, シナリオを遷移させることによって対話を実現する手法である. シナリオとは, タスクを達成するための発話手順の記述である. ここでタスクとは, シナリオの中で必要な情報を取得することである. 本研究の知識獲得におけるシナリオベース対話では, シナリオが持つタスクは, “式 3 によってあらわされた知識を獲得すること” である.

本研究における知識獲得手法では, シナリオを複数用意し, これらのシナリオを遷移させることによって, 多数の知識が獲得される. ここで, 知識を獲得したとき, その知識を用いて意味ネットワークを構築する. シナリオの遷移制御は, この意味ネットワークを用いておこなう.

3.4 獲得した知識を用いた意味ネットワーク構築

本研究では, 式 3 であらわされる知識 “ $T((b \rightarrow c)[a]) = \text{目標真理値}$ ” を獲得したとき, a を付したノード A , b を付したノード B , ならびに, c および目標真理値を付したリンク L を用いて, ノード A からノード B へリンク L が繋がった意味ネットワークを構築する.

3.5 シナリオの遷移制御

本研究では、知識獲得のインターフェースとしてシナリオベース対話を用いるため、[3]のシナリオ遷移制御とは異なる、知識獲得用のシナリオ遷移制御手法を用いる。本研究におけるシナリオ遷移制御手法は以下に示すとおりである。

1. 構築されている意味ネットワーク内で、繋がっているリンク数が最も少ないノードを検索する。
2. 1で得たノード群からランダムで1つノードを選ぶ。
3. 選ばれたノードに対して、入ってくるリンクがあった場合、そのリンクの種類によって、遷移先となるシナリオの候補を決定し、候補の中からランダムで遷移先を決定する。入ってくるリンクが無かった場合は、常に遷移可能なシナリオ候補から遷移先を決定する。
4. 1,2,3を繰り返す。

4 局所的に一貫した知識の確保

ブール多次元論理における学習とは、与えられた全ての知識において、実現値が目標値にできる限り近い値をとるようにアトムの値を決定することである。この学習において、制約条件にあたる知識の数が、ブール多次元論理の次元数に対して多いとき、間隙は小さくならない。そこで本研究は、意味ネットワークをクラスタリングによって既定数のクラスターに分割後、各クラスターごとに学習をおこなう手法を提案する。各クラスターにおいて間隙の平均を極小化することにより、知識の追加に伴う間隙の増加を抑え、局所的に一貫した知識を確保する。本章では代表的なクラスタリング手法であるk-means法と、KKZ法に倣ったクラスタリング手法および、学習の手続きについて述べる。

4.1 ハミング距離

2つの論理値を各ビットごとに比較し、異なった値をもつビットの総数を求めたとき、この総数を2つの論理値間のハミング距離とする。

4.2 提案するクラスタリング手法

KKZ法に倣った、意味ネットワークを分割するクラスタリング手法について説明する。3.4で示したとおり、本研究における意味ネットワークでは、ノードにはアトムが付され、リンクには、アトムまたは論理式、

さらに目標真理値が付されている。以降、本論文では、意味ネットワークのノードに付されたアトムを概念アトムとよび、意味ネットワークのリンクに付されたアトムを意味関係アトムとよぶ。

本研究では、概念アトムおよび意味関係アトムに対してクラスタリングを適用し、それらのアトムを各クラスターに割り当てることによって、意味ネットワークを既定数のクラスターに分割する。クラスターの分割数を k とし、概念アトムの数を n としたとき、意味ネットワークを k 個のクラスターに分割するクラスタリングの手順は以下に示すとおりである。

1. すべてのアトムと知識を対象に学習をおこない、得られたアトムの値を初期値とする。
2. 概念アトム間のハミング距離が最大となる2点を求め、その2点を初期のクラスター中心とする。
3. 全ての概念アトムに対して $D(a_i), i \in \{1, 2, \dots, n\}$ を求める。ここで $D(a_i)$ とは概念アトム a_i と既存のクラスター中心との、最短ハミング距離である。
4. 最大となる $D(a_{i'})$ の概念アトム $a_{i'}$ を次のクラスター中心とする。
5. クラスター中心を k 個選ぶまで3, 4を繰り返す。
6. 各概念アトムを、最も中心とのハミング距離が近いクラスターに割り当てる。
7. 各クラスターごとに、クラスターに所属するすべての概念アトムとの平均ハミング距離が最小となる概念アトムを求め、求めた概念アトムをクラスター中心とする。
8. クラスターの変化がなくなるまで6, 7を繰り返す。
9. 各クラスターに、全ての意味関係アトムを加える。

上記の手続きにより、概念アトムと意味関係アトムが各クラスターに割り当てられ、意味ネットワークが既定数 k のクラスターへ分割される。

4.3 各クラスターにおける学習の手順

クラスタリングによって意味ネットワークは任意に定めた分割数 k によって、 k 個のクラスターに分割される。これらの各クラスターごとに、間隙の平均を極小

化することにより、各クラスターにおける学習をおこなう。本研究における学習とは、与えられた全ての知識に対して、間隙の平均ができる限り小さくなるようにアトム値を決定することであったが、クラスターにおける学習では、クラスター C_i , $i \in \{1, 2, \dots, k\}$ の学習で用いる知識は、知識を構成するアトムが、全てクラスター C_i に含まれているもののみとする。

5 知識の一貫性が確保された範囲内での推論

本研究における推論では、論理式が与えられたとき、まず最初に、論理式を構成するすべてのアトムが同一クラスター内に含まれているかの判定をおこなう。判定の結果、対象アトムがすべて同一クラスター内に含まれていた場合は、論理式の計算をおこない結果を返す。また、判定の結果、対象アトムが同一クラスター内に含まれていなかった場合は、論理式の計算をせずに推論適用外という結果を返す。以上の手続きによってシステムは、知識の一貫した範囲内で、与えられた任意の論理式を評価し、推論をおこなう。

6 推論システムによる結果の一例

作成した推論システムの結果の一例を示す。また、本システムにおいて扱う論理値は 20 次元と定める。

6.1 知識獲得および意味ネットワーク構築部分

本システムでは 21 個のシナリオを用意し、これらのシナリオを用いたシナリオベース対話をとおして、知識の獲得および意味ネットワーク構築をおこなう。本研究では結果例として、190 個の知識群を獲得し、これらの知識にもとづいて意味ネットワークを構築した。

6.2 学習部分

提案手法により学習をおこなう。本研究では、6.1 において構築した意味ネットワークに対して 4.4 で述べたクラスタリング手法を分割数 3 において適用し、クラスタリング結果を得たのちに、各クラスターにおいて学習をおこなった。この学習の結果例を表 3 に示す

表 3. 学習をおこなった結果例

	アトム数	学習に用いた知識数	平均間隙	最大間隙
クラスタ 1	59	37	0.052	0.272
クラスタ 2	44	22	0.036	0.133
クラスタ 3	45	18	0.043	0.153

6.3 推論部分

推論部分は、任意に与えた論理式を評価する。表 3 の学習結果を用いて、論理式によってモデル化された命題に対して推論をおこなった結果の例を表 4 に示す。

表 4. 表 3 の学習結果を用いて、推論をおこなった結果の例

論理式	真理値
$((生ハム \vee 牛肉) \rightarrow 好き)[菜食主義者]$	0.666
$((弱い \vee 美味しい) \rightarrow 特徴)[魚]$	0.692
$(軍隊 \rightarrow 好き)[平和主義者]$	0.333
$(若者 \rightarrow 構成要素)[犯罪組織]$	0.800
$(猿 \rightarrow 好き)[動物愛護団体]$	0.818
$(昆虫 \rightarrow 似ている)[ミドリムシ]$	推論適用外
インテリ \rightarrow アメリカ人	0.700
システムエンジニア \rightarrow 庶民	0.750
浪人生 \rightarrow 金持ち	推論適用外

7 結論

本研究では、与えられた知識をもとに意味ネットワークを構築後、意味ネットワークを既定数のクラスターに分割し、各クラスターにおいて間隙の平均を極小化することにより、局所的に一貫した知識を確保し、知識の一貫した範囲内で推論をおこなう方式を提案した。また、シナリオベース対話を用いた知識獲得インターフェースを提案した上で、提案した方式によって構成される推論システムを設計・作成した。作成したシステムに対して、知識入力インターフェースをとおして 190 個の知識を与え、提案方式による学習および推論をおこない、その結果の一例を報告した。今後の課題としては、提案手法によってクラスタリングをおこなう際の分割数 k の決定手法の確立があげられる。また、クラスタリングを、平均間隙が指定した値を超えない範囲で、なるべく多くの知識をクラスター内に含もうとするように改良することも考えられる。

参考文献

- [1] 鈴木寿, 齋藤亮, 島田道雄, “機械知能を設計するための基本技術としてのブール多値論理の小数符号化,” 日本ロボット学会誌, vol. 22, no. 2, pp. 215–222, March, 2004
- [2] 鈴木寿, 知能情報の基礎 [C による多値論理処理], 培風館, 東京, 1999.
- [3] 大木一史, “意味ネットワークを用いたシナリオベース対話システムの研究,” 中央大学大学院理工学研究科情報工学専攻修士論文, 2009.