

記憶に基づく推論による天気予測

5J-7

—相互情報量による回答分布評価と気象庁との比較—

毛利 隆夫 田中 英彦

{mohri,tanaka}@MTL.T.u-tokyo.ac.jp

東京大学 工学部

1 はじめに

現在まで我々は、記憶に基づく推論方式 (Memory-Based Reasoning; MBR) の応用として天気予測システム WINDOM を実装し、その評価を行っている [毛利 93a, 毛利 93b]. WINDOM は、最大過去9年間 ('82~'90) 分のアメダスおよび有人観測所の観測データを保持し、マッチングにより数時間 (3~12時間) 先の関東近辺の降水の有無を予測する。「気象状況が似ているならば、その後の天気も似ているだろう」という仮定のもとに、現在の気象状況 (観測される降水や気圧などの分布) と類似した過去の時点からデータベースから検索し、その時点から数時間先の天気を、現在から同じ時間だけ先の天気予測として回答する。

本稿では、まず、実際の天気と予測した天気との相互情報量を用いて天気予測の回答分布を評価する方法を提案し、相互情報量による評価により、WINDOM による天気予測と多数派の天気しか回答しない単純な予測との比較が定量的に行えることを示す。

次に、関東甲信地区の降水有無の予測について、WINDOM による予測と、気象庁による予測との正答率での比較を行う。WINDOM は、全地域の平均値では気象庁の正答率に達していないものの、地域によっては気象庁の平均正答率に匹敵する高い正答率が得られている。

2 相互情報量による回答分布評価

問題部と回答部のペアからなる訓練事例の集合が与えられた際、それらから問題部のみが分かっている事例の回答部を予測するような、何らかの仕組みを学習する枠組みは、概念学習と呼ばれている [WK91].

一般に概念学習問題では、回答の評価は正答率を用いて行われる。しかし単に正答率だけでは、回答の分布に現れる回答の質は考慮されない。例えば、表1に挙げる3通りの回答分布を考えてみる。天気予測と対応させると、Xは正解 (実際の天気)、Yは回答 (予測された天気) を示し、天気は晴または雨の2区分とし、実際の天気の分布が、晴=80%、雨=20%である場合に相当している。

(1)~(3)とも、正答率は80%であるが、回答の分布としては大きく異なる。(1)は、常に晴と答えている。(2)は、回答の比率は、正解の比率と同じ、80:20である。(3)は、正答率80%という枠内で、雨を最も多く回答している。

(1)	<table border="1"> <tr> <td></td> <td>X</td> <td>晴</td> <td>雨</td> <td></td> </tr> <tr> <td>Y</td> <td>\</td> <td>0.8</td> <td>0.2</td> <td></td> </tr> <tr> <td>晴</td> <td></td> <td>1.0</td> <td>0.8</td> <td>0.2</td> </tr> <tr> <td>雨</td> <td></td> <td>0.0</td> <td>0.0</td> <td>0.0</td> </tr> </table>		X	晴	雨		Y	\	0.8	0.2		晴		1.0	0.8	0.2	雨		0.0	0.0	0.0	相互情報量 = 0 [bits]
	X	晴	雨																			
Y	\	0.8	0.2																			
晴		1.0	0.8	0.2																		
雨		0.0	0.0	0.0																		
(2)	<table border="1"> <tr> <td></td> <td>X</td> <td>晴</td> <td>雨</td> <td></td> </tr> <tr> <td>A</td> <td>\</td> <td>0.8</td> <td>0.2</td> <td></td> </tr> <tr> <td>晴</td> <td></td> <td>0.8</td> <td>0.7</td> <td>0.1</td> </tr> <tr> <td>雨</td> <td></td> <td>0.2</td> <td>0.1</td> <td>0.1</td> </tr> </table>		X	晴	雨		A	\	0.8	0.2		晴		0.8	0.7	0.1	雨		0.2	0.1	0.1	相互情報量 = 0.087 [bits]
	X	晴	雨																			
A	\	0.8	0.2																			
晴		0.8	0.7	0.1																		
雨		0.2	0.1	0.1																		
(3)	<table border="1"> <tr> <td></td> <td>X</td> <td>晴</td> <td>雨</td> <td></td> </tr> <tr> <td>Y</td> <td>\</td> <td>0.8</td> <td>0.2</td> <td></td> </tr> <tr> <td>晴</td> <td></td> <td>0.6</td> <td>0.6</td> <td>0.0</td> </tr> <tr> <td>雨</td> <td></td> <td>0.4</td> <td>0.2</td> <td>0.2</td> </tr> </table>		X	晴	雨		Y	\	0.8	0.2		晴		0.6	0.6	0.0	雨		0.4	0.2	0.2	相互情報量 = 0.322 [bits]
	X	晴	雨																			
Y	\	0.8	0.2																			
晴		0.6	0.6	0.0																		
雨		0.4	0.2	0.2																		

表1: 回答分布と相互情報量 (正答率はいずれも80%)

(1)~(3)の回答は、正答率を用いれば同一視されるが、相互情報量を利用すれば容易に区別できる。ここで、X、Yの各事象を $x_i, y_j (i, j = 1 \dots N)$, その確率を $p(x_i), p(y_j)$, x_i と y_j が同時に起こる確率を $p(x_i, y_j)$ とする。また、 $H(X)$ をXのエントロピー、 $H(X|Y)$ をYを知ったもとでのXのエントロピーだとすると、確率事象系XとYの相互情報量 $I(X;Y)$ は次のように定義できる。なお、 $I(X;Y)$ は、 $0 \leq I(X;Y) \leq \min\{H(X), H(Y)\}$ の値をとり、XとYが独立の時に0になる。

$$I(X;Y) = \sum_i^N \sum_j^N p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

$$= H(X) - H(X|Y)$$

天気予測の場合には、 $I(X;Y)$ は、予測Yを知った時に、実際の天気Xのエントロピーが減少する割合を示している。つまり、予測を知った後で、どれだけ実際の天気を特定しやすくなったかが、相互情報量で定量的に表現されている。

相互情報量を用いた場合、表1の3通りの予測の相互情報量は、それぞれ表右側のようになる。特に、(1)のような単一の項目しか回答しない場合は、正解と回答が独立しているため、相互情報量が0 bitsになり、他の場合と明確に区別される。(2)と(3)を比較した場合、(3)のような予測の方が、より実際の天気を特定しやすい予測であることも分かる。

回答の分布を評価する他の方法としては、コスト関数を用いる方法があるが、回答の種類毎のコストが予測の利用者に变化し、値の設定に恣意性が伴う。一方、相互情報量による評価は、利用者のモデルを仮定するなどの

⁰Weather Prediction Using Memory-Based Reasoning:
- Evaluation of the Distribution of Replies by Mutual Information and Comparison with the Meteorological Agency -
Takao MOHRI and Hidehiko TANAKA
Faculty of Engineering, The University of Tokyo

必要はなく、その理論的根拠も明確である。

ただ、相互情報量による評価では、完全に嘘を答えた場合にも高い値が得られる。例えば、実際の天気とは完全に逆に晴と雨を予測したとしても、最大の相互情報量が得られる。このように、相互情報量の上昇は、必ずしも正答率の上昇を意味しない点に注意する必要がある。したがって、正答率を主な評価尺度として考え、相互情報量は副次的に利用するのがよいと考えられる。

3 相互情報量による天気予測の評価

相互情報量による天気予測の評価を、我々が行っている MBR による天気予測に適用した。予測時間(何時間先の天気を予測するか)と、東京大手町一点での降水有無予測の正答率との関係を図 1 に示す。ここで、major-weather は、最も多数を占める回答カテゴリを回答する単純な方法である。東京の天気は、約 84% が降水無し(晴)、残りが降水あり(雨)であるため、major-weather は、必ず「晴」を回答する。

図 1 では、12 時間先の予測では、WINDOM と major-weather との差が小さくなっているが、回答の分布は大きく異なっている。実際、異なる時刻に対して 1000 回行ったテストにおいて、12 時間後に降水がある場合は 167 回あったが、WINDOM はそのうち 61 回を正答しているのに対して、major-weather は 1 回も正答していない。

相互情報量による回答分布の評価を行ったのが図 2 であり、両者は明確に区別される。major-weather は、必ず同じ回答を行うため相互情報量は必ずゼロであるが、WINDOM は情報のある予測を行っていることが分かる。

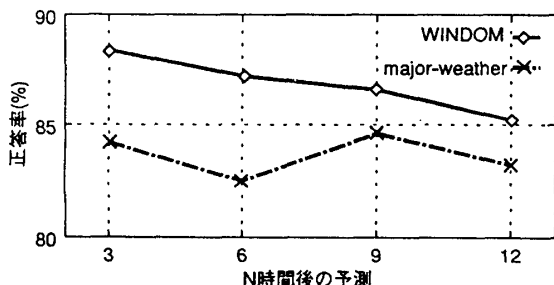


図 1: 予測時間と正答率の関係

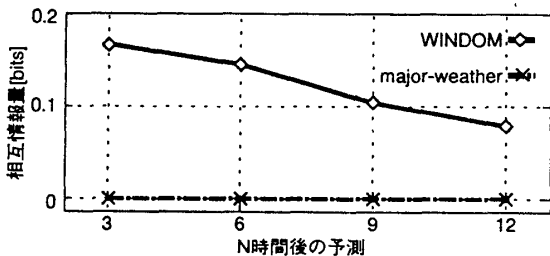


図 2: 予測時間と相互情報量との関係

4 気象庁の予測結果との比較

現在、気象庁では日常的に天気予測が行われており、一部の予測の正答率が公表されている [気象 92]。降水有無の予測に関して正答率が公表されている区域で、東京を含む最小のものは、関東甲信地方 (関東 1 都 6 県 + 山梨 +

長野) である。本実験では、1991 年の毎朝の朝予報 (6 時から 18 時までの 12 時間予報) を用いて、気象庁の予測と WINDOM とを比較した。

WINDOM は特定の地点の固定された時間先の予測を行う方法であるが、気象庁の朝予報は地域 (各都県の約半分の区域) を対象にし、時間の幅も持っている。ここでは簡単に、地域内の一地点の 6 時間先の予測を、その地域の全地点での 1~12 時間先の予測として代用した。結果を表 2 に示す。正答率は、平均値では気象庁と比べて、幾分劣っているが、東京都、神奈川県などは気象庁の平均正答率に近い値が得られている。なお長野県の正答率は他都県に比べて大きく下回っているが、原因の一つとして、長野県 (実際は北部) が他の都県と比べて区域内での天気のバラツキが大きく、区域単位での予測では高い正答率が得づらい点が挙げられる。

表 2: 気象庁による予測の正答率との比較

方法	予測区域	正答率 [%]
WINDOM	東京都	86.3
	神奈川県	87.9
	埼玉県	84.6
	千葉県	84.0
	群馬県	83.2
	栃木県	86.1
	茨城県	82.7
	山梨県	83.7
	長野県	74.4
	関東甲信平均	83.7
気象庁	関東甲信平均	87

5 結論および今後の課題

本稿では、2 つの方法によって我々の天気予測システム WINDOM の評価を行った。まず、概念獲得問題の回答分布を相互情報量により評価する方法を提案し、その特徴を論じた後、WINDOM の予測結果を相互情報量を用いて評価し、単純に多数派の天気を回答する方法とは明確に区別できることを示した。次に、気象庁の行っている予測と WINDOM との正答率での比較を行い、関東甲信地方の 1~12 時間先の降水有無予測について、平均正答率では及ばないものの、地域によってはそれに匹敵する高い正答率が得られていることがわかり、本手法の有効性が実証できた。

今後の課題としては、より長時間先の予報への対応や、高層気象などのデータの追加利用による正答率向上が挙げられる。

参考文献

[WK91] Sholom M. Weiss and Casimir A. Kulikowski. *Computer Systems That Learn*. Morgan Kaufmann, 1991.

[気象 92] 気象庁天気相談所. 降水の有無予報及び最高・最低気温予報の評価結果 (1991 年), 1992.

[毛利 93a] 毛利隆夫, 中村真章, 田中英彦. 記憶に基づく推論方式による天気予報. 人工知能学会全国大会 (第 7 回) 論文集, pp. 43-46, 1993.

[毛利 93b] 毛利隆夫, 田中英彦. 記憶に基づく推論方式による天気予報 - システムの性能評価 -. 第 47 回情報処理学会全国大会講演論文集 (2), pp. 41-42, 1993.