

AI の確率はどんな価値を持つか

How valuable are probabilities produced by AI?

中村國則 木村元優[†]

Kuninori Nakamura, Motohiro Kimura

[†] 成城大学社会イノベーション学部

Faculty of Social Innovation, Seijo University

knaka@seijo.ac.jp

概要

本研究は、AI が与える情報と人間の専門家が与える情報の解釈の相違を、確率値の情報としての価値に対する判断という側面から検討した。実験では、参加者は“企業の株価が上昇する見込み”に関する情報として様々な確率値を与えられたうえで、それらの確率値が情報としてどの程度参考になるかを、情報がAI から与えられた場合と人間の専門家から与えられた場合のいずれかで評価することを求められた。分析の結果、AI の与える情報を解釈する際、参加者は人間の専門家が与える場合よりも低い事前確率を見込んでいること、情報自体の影響も弱いことが明らかになった。

キーワード：AI, 確率, 情報量

1. はじめに

AI や統計モデルからの助言を人がどう理解するか、という問題は1950年代(e.g., Meehl, 1954)から近年(e.g., Dietvorst et al, 2015, Logg et al, 2019)に至るまで様々な検討が加えられてきた。そこでは、AI や統計モデルの判断よりも人間の専門家の判断の方を信頼するというアルゴリズム忌避(Dietvorst et al, 2015)と、逆にアルゴリズムの判断をより信頼するというアルゴリズム選好(e.g., Logg et al, 2019)の双方が報告され、AI からの情報の受容過程が人間から与えられる場合のそれと大きく異なることが一つの心理学的事実として確立されてきた。近年、日常生活の様々な側面でAI 利用が進みつつある中、このようなアルゴリズム忌避・選好が人間の意思決定支援場面でどのように生じているかは重要な検討課題といえるだろう。

その一つの局面として考えられるのが、AI がもたらす確率情報のような“言葉”を人がどう理解するか、という問題である。これまでの研究から、人は確率という情報の価値を値そのものの大小(Keren & Teigen, 2001)や事前の信念との対比(中村, 2008)から評価していることが知られている。では、そもそもこの確率という情報に対する評価の仕方は、それがAI がもたらしたものか人間がもたらしたものかで異なるのだろうか。上で述べたアルゴリズム忌避・あるいは選好を示した先行研究を踏まえれば、両者の間で違いが生じることは十分

考えられる。本研究の目的は、このような相違があるかを検討することである。

本研究のもう一つの目的は、このようなAI と専門家の相違を、先行研究(Keren & Teigen, 2001; 中村, 2008)で確認された、確率の価値に対する情報理論的な解釈から分析することである。Keren & Teigen (2001)によれば、人は確率という情報の価値づけについて、中程度の確率よりは極端に低い・あるいは高い確率を、そして低い確率よりは高い確率をより情報として有益とみなす“明確な予測の探求原理”とでもいえるべき基準で評価しているとされる。そして中村(2008)は、このような傾向を、確率という情報の価値は、物事の起こりやすさに関する事前の主体が有する信念(q)と確率値(p)で表されるKullback-Leibuller 情報量、

$$D_{KL}(P||Q) = \sum_i P_i \log \frac{Q_i}{P_i} \quad (1)$$

に従い、かつ多く場合物事は起こりにくい($q < 0.5$)という仮定の下で情報量を見積もるためであることを示した。この解釈に従えば、AI に基づく情報をどのような事前信念に基づいて解釈しているのか、そしてその事前信念は人間が与える情報とどのように異なるのかを検討することが可能である。同時に、AI が与える確率情報という状況の下で、先行研究(Keren & Teigen, 2001; 中村, 2008)で確認された傾向が再現できるのか、という問題を検討することにもなるだろう。

以上の目的の元、本研究では、“ある企業の株価の上昇する見込み(Mislavsky & Gaertig, 2022)”という場面を用いて、その上昇する見込みを表す確率値に対する情報の価値に対する参加者の評価を、上昇する見込みを述べたのがAI である状況と人間の専門家である状況との間で比較した。

2. 方法

222名の私立大学生が授業の一環として実験に参加し、実験の説明・およびデータの収集は全てGoogle formを通じて行った。回答者は図1の画像である企業のそ

銘柄コード	企業名	株価
GIL	ギルダン・アクティブウェア	\$31.84

図 1 実験参加者に与えられた企業の架空の株価

表 1 本研究で用いた言語表現	
ポジティブ表現	ネガティブ表現
かすかな見込み	ない
僅かな見込み	ほぼ不可能
少しだけ考えられる	あまりない
多少ありそうである	それほどでない
ある	確信がない
考えられる	やや疑わしい
十分に考えられる	確実ではない
確実である	多少の不安がある

の時点での株価を示された上で、その企業の株価が次週に上がる見込みを表す確率情報についてそれがどの程度情報として参考になるかを回答することを求められた。確率情報として 7 種類の確率値(1%, 17%, 33%, 50%, 67%, 83%, 99%), および 16 種類の言語確率表現(表 1 を参照)を提示した。これら合計 23 種類の確率表現について、120 名の参加者はこれらの情報が AI からもたらされた場合について、102 名の参加者は株の専門家からもたらされた場合のどちらかの状況の下で、どの程度参考になるかを 5 段階(“全く参考にならない”: 1~ “非常に参考になる”: 5)で回答する。また、16 種類の言語確率表現については個々の表現が示す確率値をパーセント単位で回答した。

3. 結果および考察

専門家条件・AI 条件の平均評定値を図 2 に示す。ここで、中村(2008)の想定に基づき、情報の有益さに対する参加者の判断は、事前の信念に基づいて評価された確率情報の情報量に対応すると仮定し、参加者の評定値に対して以下のモデルをあてはめ、事前の信念を推定した；

$$y = a + b \times \left\{ p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q} \right\} + c_1 d_{num} + c_2 d_{negative} \tag{2}$$

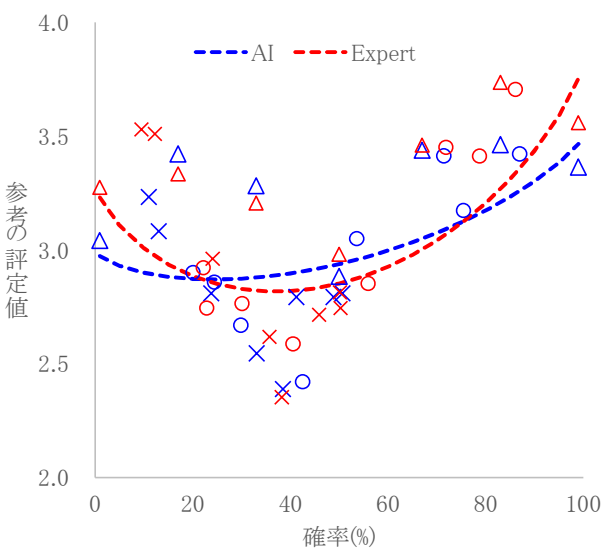


図 2 AI 条件・専門家条件における平均評定値とモデルの分析結果:図中の曲線は非線形混合モデルで推定された q , a の値に基づく。

ここで y は参加者の判断、 a , b は切片、情報量に対する係数、 c_1 , c_2 は方向性をポジティブ表現を基準としたダミー変数として表した際の、数値確率、ポジティブ表現に対する係数を表し、 p は提示された確率情報が表す確率値、 q は事象の起こりやすさに対する事前の信念である。このモデルでは、与えられた確率情報の有益さ(y)は、与えられた確率情報が表す確率値(p)の Kullback-Leibuller 情報量を、参加者が想定する事前の事象の起こりやすさに対する信念(q)に基づいて評価することを仮定し、データから参加者の事前の信念やその信念に対する判断の重みづけ、方向性の影響を推定するものである。パラメタ推定は非線形混合モデルに基づいて行い、 q , c_1 , c_2 については固定効果を仮定し、 a , b についてはランダム効果を仮定した。また、 p については数値確率については提示した確率の値を、言語確率については参加者の評定した確率値を代入した。

推定結果を表 2 に示す。AI 条件・専門家条件ともに、推定された q の値は 50%を下回っており、先行研究(中村, 2008; Keren & Teigen, 2001)の知見を再現した。また、専門家条件と比較して AI 条件の方が推定された q の値は低く、かつ KL 情報量の項に対する重みづけも低かった。また、言語確率の方向性の影響については AI 条件、専門家条件ともに数値確率がポジティブ表現よりも、ポジティブ表現がネガティブ表現よりも高く有益さを見積もられていた。

表 2 モデルの推定結果：カッコ内は 95%信頼区間を示す。

パラメタ	条件	
	専門家	AI
a	2.82	2.86
	(2.69, 2.94)	(2.76, 2.98)
b	1.00	0.44
	(0.79 ,1.21)	(0.43 ,0.45)
c1	0.24	0.23
	(0.14, 0.35)	(0.13, 0.33)
c2	-0.11	-0.14
	(-0.20, 0.00)	(-0.24, -0.05)
q	0.37	0.24
	(0.32, 0.42)	(0.14, 0.34)

このような知見は、AI の与える情報については株価の上昇については専門家の場合と比較して相対的に“起こりにくい”という事前信念の下で情報を見積もると同時に、確率値自体の情報量の影響も低いことを示している。発表当日はこの結果の理論的意義についてより詳細な検討と議論を試みたい、

4. 引用文献

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114-126.

Keren, G., & Teigen, K. H. (2001). Why $isp = .90$ better than $p = .70$? Preference for definitive predictions by lay consumers of probability judgments. *Psychonomic Bulletin and Review*, 8(2), 191-202.

Meehl, P. E. (1954). *Clinical versus statistical prediction*. Minneapolis, MN: University of Minnesota Press.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103.

中村國則 (2008) "十分にありえる"方が"見込みがない"より意味があるか? : 言語確率の情報量の定量的

分析とその情報理論的解釈. *認知科学*, 15, pp.174-187.