Analogies and Theories: The Role of Simplicity and the Emergence of Norms

Gabi Gayer and Itzhak Gilboa

May 16, 2013

Modes of Reasoning

- Rule-based

    Theories

- Case-based

    Analogies; similarities

- When will one type of reasoning be more common than the other?

Environment

- Exogenous: The outcome is independent of the prediction- weather
- Endogenous: The outcome is determined by the prediction- social norms, equilibrium
- A combination of the two possibilities above- stock market

Framework

- Adapt the model of induction of GSS (2010)
- Assumptions about how y is determined

General Model

The primitives are:

- X – a set of characteristics
- Y – a set of outcomes

$$0 < |X|, |Y| < \infty$$

- $\Omega = (X \times Y)^{\infty}$ – the set of states of the world
- $\omega(t) = (\omega_X(t), \omega_Y(t))$-the element of $X \times Y$ in period t given state ω

Some More Notation

- For a state $\omega \in \Omega$ and a period t, the a history up to period t

$$h_t = (\omega(0), \ldots, \omega(t-1), \omega_x(t))$$

  and its associated event

$$[h_t] = \{\omega \in \Omega \mid (\omega(0), \ldots, \omega(t-1), \omega_x(t)) = h_t\}$$

- For a history h_t and a subset of outcomes $Y' \subset Y$ define the event

$$[h_t, Y'] = \{\omega \in [h_t] \mid \omega_y(t) \subset Y'\}$$

  namely, the event in which history h_t occurs followed by an outcome in Y'.

Predictions

- At each period $t \in T$, the agent observes a history h_t and makes a prediction about the period-t outcome, $\omega_Y(t) \in Y$.

- More generally, a prediction is a ranking of subsets in Y given h_t
- Predictions are made with the help of conjectures. A conjecture is an event $A \subset \Omega$.
- The set $\mathcal{A} \subset 2^{\Omega}$ – the set of permissible conjectures CB, RB

The Weighting Function

- A function $\phi : A \rightarrow R_+$, $\phi(A)$ is the weight (or degree of belief) the agent attaches to conjecture A
- For a subset of conjectures $D \subset A$ inference is driven in an additive manner

$$\phi(\mathcal{D}) = \sum_{A \in \mathcal{D}} \phi(A)$$

- Convention: $\phi(A) = 1$

Reasoning by Conjectures

- Reasoning or learning is pseudo-Bayesian - conjectures that were proven to be inconsistent with h_t are discarded
  Given history h_t, all conjectures A such that
  $$A \cap [h_t] = \varnothing$$
  are refuted and should be discarded.
- Conjectures A such that
  $$A \cap [h_t] = [h_t, Y]$$
  say nothing and are irrelevant.

How Likely is a Set of Outcomes?

- Given history h_t, how much credence does ϕ lend to each outcome? Or to each set of outcomes?
- For $Y' \subsetneq Y$ define

$$\mathcal{A}(h_t, Y') = \left\{ A \in \mathcal{A} \,\middle|\, \varnothing \neq A \cap [h_t] \subset \left[h_t, Y'\right] \right\}$$

which is the class of conjectures that:

(1) have not been refuted by h_t
(2) predict that the outcome will be in Y' (hence relevant)

- Their weight

$$\phi(\mathcal{A}(h_t, Y')) = \sum_{A \in \mathcal{A}(h_t, Y')} \phi(A)$$

is the degree of support for the claim that the next observation will be in Y'.

More Notation

For a subset $D \subset A$ define the set of conjectures in D that have nor been refuted and predict an outcome in $Y' \subsetneqq Y$

$$\mathcal{D}(h_t, Y') = \left\{ A \in \mathcal{D} \,\middle|\, \varnothing \neq A \cap [h_t] \subset \left[h_t, Y'\right] \right\}$$

- Also the total weight of all conjectures in D that have not been refuted and are relevant:

$$\phi(\mathcal{D}(h_t)) = \phi\left(\cup_{Y' \subsetneqq Y} \mathcal{D}(h_t, Y')\right)$$

- We have a special interest in subsets of conjectures: RB, CB.

Rule-Based Reasoning

- Make a specific prediction (i.e., a single y_t) at each and every t, and for any possible value of x_t.
- The set of rule-based conjectures is denoted by R is assumed countable
- ... and to contain a theory for every possible history
- For example, the set of computable theories
- Let $\phi(f_j) = \phi([f_j])$ be the weight assigned by ϕ to theory f_j.
- A model ϕ is purely rule-based if all weight is put on rule-based conjectures.

Special Case: Bayesian

Special case: no x values (that is, $|X| = 1$).
Then for every $f_j \in R$, there is a unique state compatible with it.
R can also be viewed as a Bayesian model (as defined in GSS, 2013) that assigns probabilities to single states.

Case-Based Reasoning

- Consider a simple case-based model of prediction. For a similarity function

$$s : X \times X \rightarrow \mathbb{R}_{+}$$

define the aggregate similarity for an outcome $y \in Y$

$$S(h_t, y) = \sum_{i=0}^{t-1} \beta^{t-i} s(\omega_x(i), \omega_x(t)) \mathbf{1}_{\{\omega_y(i)=y\}}$$

- This is equivalent to kernel classification (with the similarity playing the role of the kernel).

Case-Based cont.

- The case-based conjectures will be of the form

$$A_{i,t,x,z} = \{\omega \in \Omega \,|\omega_x(i) = x, \omega_x(t) = z, \omega_y(i) = \omega_y(t)\}$$

  for periods $i < t$ and two characteristics $x, z \in X$.
- $A_{i,t,x,z}$ can be viewed as predicting:
  "in period i we'll observe characteristics x, in period t we'll observe characteristics z, and the outcomes will be identical"
- Or:
  "IF we observe characteristics x and z in periods i and t, (resp.) THEN we'll observe the same outcomes in these periods."

Case-Based cont.

- The set of all case-based conjectures is

$$\mathcal{CB} = \{ A_{i,t,x,z} \mid i < t\,,\ x, z \in X \} \subset \mathcal{A}.$$

- To embed a similarity model, with $s : X \times X \to R_+$ in the GSS model, define

$$\phi_{s,\beta}(\{A_{i,t,x,z}\}) = \beta^{(t-i)} s(x,z)$$

  to get

$$S(h_t, y) = \phi_{s,\beta}(\mathcal{A}(h_t, \{y\}))$$

- A model ϕ is purely case-based when all weight is put on case-based conjectures.

Special Case: Frequentist

When $\beta = 1$ and $s(x, z) = c$

$S(h_t, y)$ is proportional to the empirical frequency of y's in the history h_t.

Assumptions

- $A = R \cup CB$
- Open-Mindedness: $\phi(A) > 0$ for each conjecture in $A = R \cup CB$. Denoted by Φ_+.

Exogenous Process: Simplicity Result

- All states $\omega \in [f_j]$ are simple (say, describably by a Turing machine)

Proposition

For every $\phi \in \Phi_+$ *and every* $\omega \in \mathbb{S}$,

$$\frac{\phi\left(\mathcal{CB}(h_t\left(\omega\right))\right)}{\phi\left(\mathcal{R}(h_t\left(\omega\right))\right)} \to 0$$

- The idea: the initial weight assigned to this theory $\phi(f_j)$ is a lower bound of the weight of the rule-based conjectures.
- The total weight of the set of all (relevant) case-based conjectures at time t converges to zero.

Insufficiency of Rule-Based Reasoning

- Defining

$$\Omega_{R\phi} = \{\omega \in \Omega \mid \exists T, \quad \phi\left(\mathcal{R}(h_t(\omega))\right) > \phi\left(\mathcal{CB}(h_t(\omega))\right) \quad \forall t \geq T\}$$

$$\Omega_{C\phi} = \{\omega \in \Omega \mid \exists T, \quad \phi\left(\mathcal{R}(h_t(\omega))\right) < \phi\left(\mathcal{CB}(h_t(\omega))\right) \quad \forall t \geq T\}$$

$$\Omega_{M\phi} = \left\{ \omega \in \Omega \;\middle|\; \begin{array}{c} \forall T, \quad \exists t, t' \geq T, \quad \text{such that} \\ \phi\left(\mathcal{R}(h_t(\omega))\right) \geq \phi\left(\mathcal{CB}(h_t(\omega))\right) \\ \phi\left(\mathcal{R}(h_{t'}(\omega))\right) \leq \phi\left(\mathcal{CB}(h_{t'}(\omega))\right) \end{array} \right\}$$

Proposition

Let there be given a model $\phi \in \Phi_+^p$. Then $\Omega_{R\phi}$, $\Omega_{C\phi}$, and $\Omega_{M\phi}$ are dense in Ω.

Insufficiency of Rule-Based Reasoning cont.

- Most states of the world are not simple.
- Let λ be the uniform probability measure: $\lambda([h_t]) = \lambda([h'_t])$

Insufficiency of Rule-Based Reasoning cont.

- What about the decay rate of case-based conjectures?

Assuming $\Phi_+^p \subset \Phi_+$ to be the set of models ϕ for which there exist $\gamma < -1$ and $c > 0$, such that, for every t, and every $x, z \in X$,

$$\sum_{i<t} \phi\left(A_{i,t,x,z}\right) \geq ct^{\gamma}$$

Proposition

For every $\phi \in \Phi_+^p$

$$\lambda\left(\frac{\phi\left(\mathcal{R}(h_t\left(\omega\right))\right)}{\phi\left(\mathcal{CB}(h_t\left(\omega\right))\right)} \rightarrow_{t\rightarrow\infty} 0\right) = 1.$$

Endogenous Process

- Assumption: all agents share the same weight function $\phi \in \Phi$.
- The outcome is determined by the agents' prediction:$\Omega_\phi = \{\omega \in \Omega \mid \omega_Y(t) \in \arg\max_{y \in Y} \phi(\mathcal{A}(h_t, \{y\})) \quad \forall t \geq 0\}$

Proposition

For every $\omega \in \Omega$, there exists $\phi \in \Phi_+^p$ such that $\omega \in \Omega_\phi$.

- The proof is constructive describing an algorithm finds ϕ that generates ω.
- But ϕ may also not be computable (if ω is not computable)
- Restrict attention to models ϕ that are computable $\Phi_+^{cp} \subset \Phi_+^p$, such that, for each $\phi \in \Phi_+^{cp}$, there is a machine that computes $\phi(A) \in Q$ for each conjecture A.

Endogenous Process- Domination of Rule-Based vs. Case-Based Reasoning

- Rule-based reasoning is dominant at state $\omega \in \Omega_\phi$ at period t if

$$\begin{aligned} (i) \quad \phi\left(\mathcal{R}(h_t\left(\omega\right))\right) &> \phi\left(\mathcal{CB}(h_t\left(\omega\right))\right) \\ &\textit{and} \\ (ii) \quad \omega_Y\left(t\right) &\in \arg\max_{y\in Y} \phi(\mathcal{R}(h_t, \{y\})). \end{aligned}$$

- Case-based reasoning is dominant at state $\omega \in \Omega_\phi$ at period t if

$$\begin{aligned} (i) \quad \phi\left(\mathcal{R}(h_t\left(\omega\right))\right) &< \phi\left(\mathcal{CB}(h_t\left(\omega\right))\right) \\ &\textit{and} \\ (ii) \quad \omega_Y\left(t\right) &\in \arg\max_{y\in Y} \phi(\mathcal{CB}(h_t, \{y\})). \end{aligned}$$

- It could by that neither is dominant

Endogenous Process-Results

- $\Omega_{RB\phi}$ – the set of states $\omega \in \Omega_\phi$ such that for some T, rule-based reasoning is dominant at state $\omega \in \Omega_\phi$ at all $t \geq T$.
- $\Omega_{CB\phi}$ – the set of states $\omega \in \Omega_\phi$ such that for some T, case-based reasoning is dominant at state $\omega \in \Omega_\phi$ at all $t \geq T$.

Proposition

For every $\phi \in \Phi_+^{cp}$ we have $\mathbb{S} \cap \Omega_\phi \subset \Omega_{RB\phi}$.

Proposition

For every $\phi \in \Phi_+^{cp}$ we have $\Omega_{CB\phi} = \varnothing$.

- But $\Omega_{RB\phi} \neq \Omega_\phi$. For example it is possible that the agent's reasoning will fluctuate between rule-based and case-based reasoning.

Hybrid Model

- The $y_t - s$ are governed partly by the predictions $\hat{y}_t$ (endogenous) and partly by random shocks (exogenous)

$$y_t = \begin{cases} \hat{y}_t & \text{with probability } \alpha\left(h_t\right) \\ \tilde{y}_t & \text{with probability } 1 - \alpha\left(h_t\right) \end{cases}$$

where $\hat{y}_t \in \arg\max_{y \in Y} \phi(A(h_t, \{y\}))$ and $\tilde{y}_t$ is uniformly distributed.

- If α does not depend on history then for every $\alpha \in (0, 1) \longrightarrow$ tends to be case-based
- If α depends on history - for example converges to 1 with t when a rule is not refuted $\longrightarrow$ tends to be rule-based

Conclusions

- Exogenous Process $\longrightarrow$ case-based
- Endogenous Process $\longrightarrow$ rule-based
- Housing Market:
  Rent: more case-based
  Sales: more rule-based
- Heterogenous beliefs