AGI engineering

Subscribe

DECISION-MAKING FUNDAMENTALS

SCIENCE AND ART OF COMPROMISES

Mykola Rabchevskiy

May 9 ♡ 💬 4 ⤴

The previous chapter, Nuts and bolts of the decision making, discussed the specificity of the decision-making process associated with the continuity of the material environment. This chapter deals with those fundamental features of decision-making that always take place regardless of whether the environment is continual material or discrete virtual.

The decision about what actions should be taken is somehow based on the predictions of the consequences of possible actions. Forecasts, in turn, can be based both on individual experience and on knowledge gained from outside. Using only one's own experience is the most challenging situation, confirmed by both human experience and experiments with AI systems. Since learning from experience itself is one of the main differences between AGI and narrow AI, the following analysis discusses aspects of decision-making based on individual experience.

The more distant consequences of the actions analyzed by the system, the more successful its mission will be. This means that making intelligent decisions requires predicting the outcomes of chains of sequential actions.

It is reasonable to consider regularly repeated sequences of actions as new actions (*composite*, as opposed to elementary *atomic* actions), replenishing the set of possible

📑 Publish on Substack

AGI engineering is on Substack – the place for independent writing

An elementary forecast includes a few new situations that, based on individual experience, are possible after performing a particular action. Putting together the forecasts for all allowed actions in the current situation, we get a complete one-step forecast. Repeating the forecasting process for each of the possible situations after the first step, we will expand the *forecast horizon* to two stages and further. The diagram illustrates the principle (actions are designated by the letters *A, B, ..*, gray color means no experience for the combination of situation+action, red means prohibition of action in this situation):
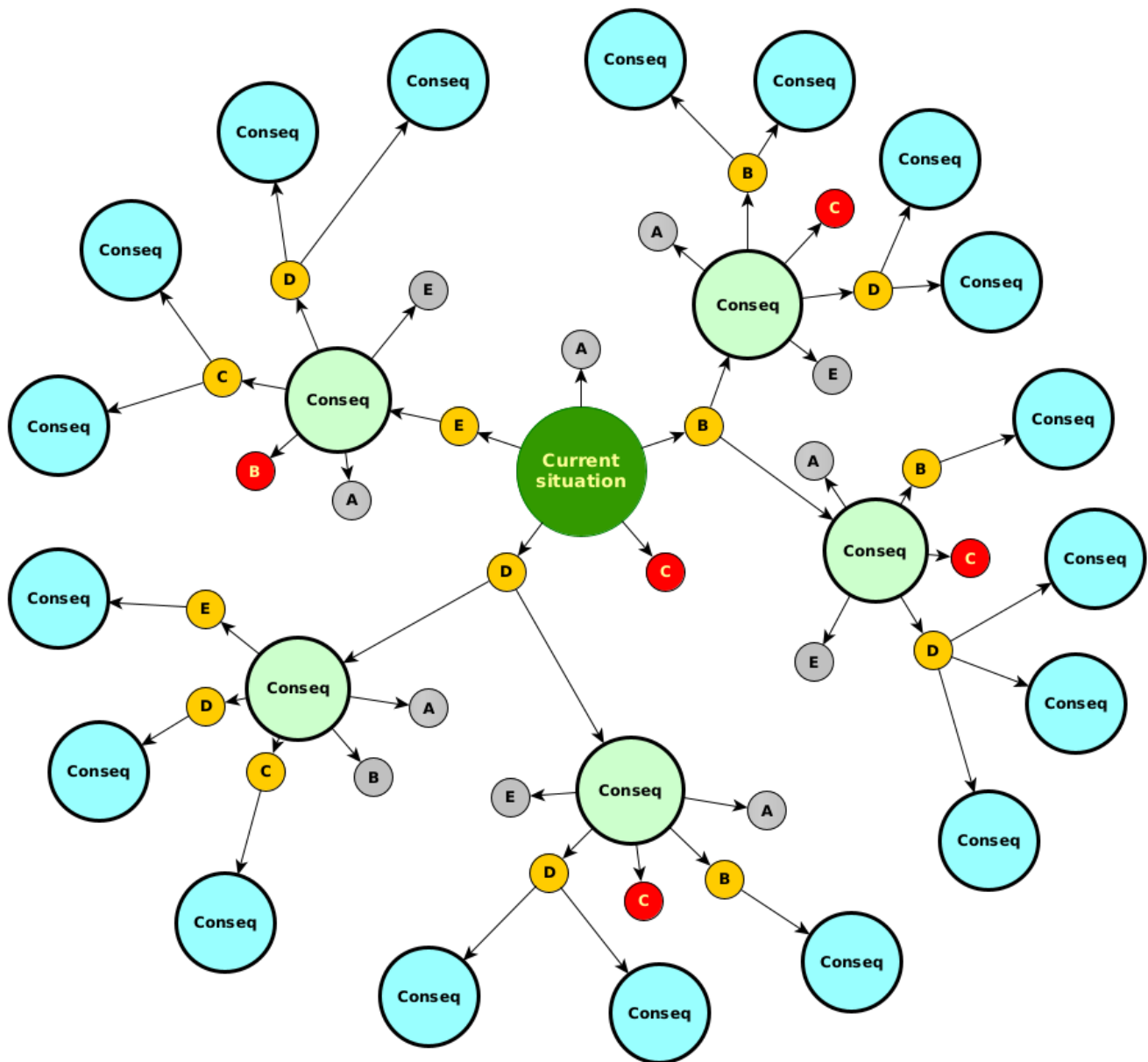
In the process of building a multi-stage forecast, several potential problems are apparent. The number of possible consequences can grow catastrophically quickly in the case of many potential actions, limiting the forecast horizon due to the limited available memory and time - the forecast must be used before it becomes outdated. For many combinations of situations and actions, individual experience may be lacking and thus

limit the forecast horizon (in the case of multi-stage prognosis, this is, to some extent, a "counterbalance" to the factor mentioned above). Finally, as the forecast horizon increases, there is a tendency to obtain, from some point on, an "anything is possible" forecast that is of no practical value.

Here is the time to remember that actions are understood as atomic actions or sequences of atomic actions that have received the status of the action. This means that the forecast horizon, measured by the length of the chain of atomic actions, can grow as experience is gained and composite actions are created.

In the case of a continuous environment, actions have a duration measured in real-time, and the forecast horizon is also measured by an interval of time.

Reducing the resource intensity of forecasting is facilitated by the fact that different chains of actions can lead to identical situations (this requires a quick way to compare situations for identity). In addition, if a specific situation is certainly unacceptable, there is no point in evaluating further actions.

This is the perfect time to move from discussing the construction of forecasts to what it is made for - to evaluations of situations that potentially arise due to specific actions. The ultimate goal of building a prediction is the optimal choice of the action (or inaction) that should be implemented in a given situation. As a result of making a forecast, there is a set of possible situations and a chain of actions from the current situation to each possible outcome. For each final state, we can analyze the chains of events in reverse order and find one or more paths, which makes it possible to achieve this final state.

The choice of the preferred final state implies the ability to *sort* potentially accessible situations according to the degree of preference, that is, the presence of a *function that calculates the situations' rating* corresponding to the current motivation. In the same way, we can evaluate each intermediate situation in the chain from the current state to the final one. Further, the *whole path can be characterized by a pair of ratings: the minimal rating for the entire route and the rating of the final state*.

Obviously, from the point of view of the system's mission, it is not only the final situation rating that matters but also how bad/good the intermediate situations are.

Combining the paths starting with a specific initial action, we get a pair of ratings characterizing the initial action: the rating of the least desirable state that is possible if this action is chosen, and the highest rating that is ultimately achievable when this action is selected. So *each of the actions allowed in the initial state now has two criteria for*

making a decision: a rating of the best possible result B (best) and a rating of the worst possible situation W (worst) within the forecast horizon.

Accordingly, each possible action *i*, for which there is an individual experience, is represented by a point on the plane in coordinates (*Wi, Bi*). This reduces decision-making to a *two-criterion optimization problem*:
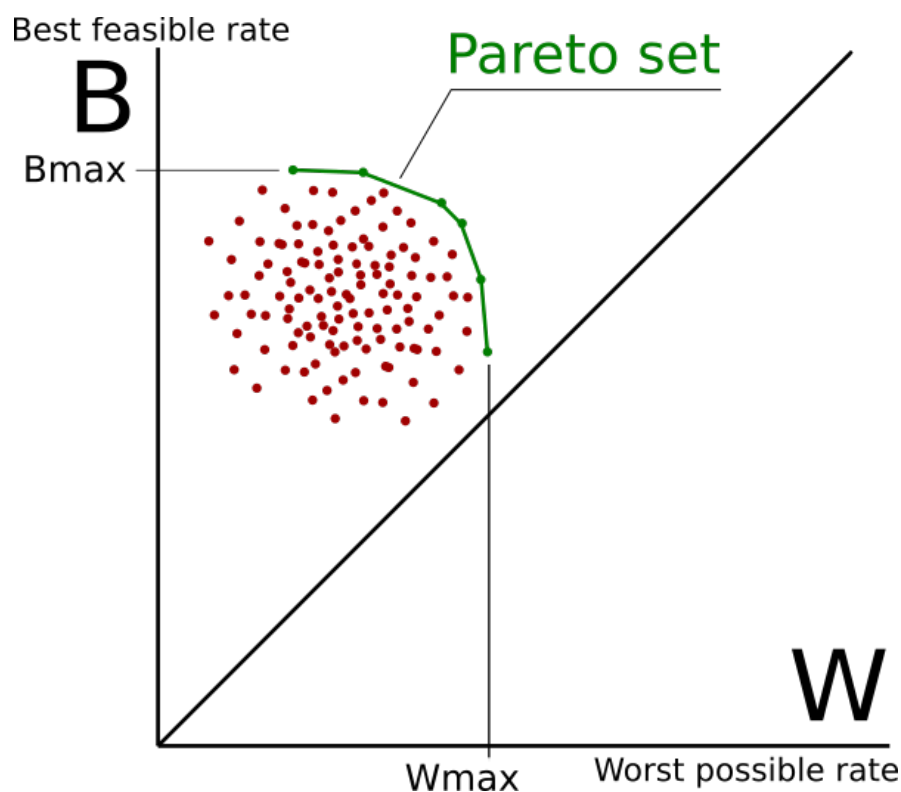
There are two extreme cases of choice. We can choose an action based on maximal **B** regardless of the risk of getting into an undesirable situation. Alternatively, we can choose an action based on maximal **W** - to reduce the risk of undesirable situations, regardless of the final rating. Naturally, in the general case, there are intermediate compromise options as a result of *maximization of the weighted sum of two criteria*:

action(k) = arg max (k*Bi + (1-k)*Wi)

The set of actions corresponding to different **k** in the range [**0, 1**] constitutes a *Pareto set*, a *set of compromise solutions for which it is impossible to improve one of the optimality criteria without worsening the other*.

This circumstance is of fundamental importance: the choice of the optimal value requires the assignment of the value k, which determines the point of compromise between a "greedy" strategy and a cautious one. The assignment of the value **k** (and, accordingly, the choice of one of the compromise options for actions) *cannot be the result of a formal analysis*. There is no single best solution; the compromise between benefits

and risk is chosen by the system designer either explicitly or indirectly through the rules implemented by the motivation module. *Implementing intelligent decision-making requires combining the science of optimization with the art of the system designer in choosing the point of compromise*.

Varying the compromise point by the motivation module depending on the circumstances allows rationalizing the assignment, for example, using risky behavior during the period of knowledge/skills accumulation in a safe "learning" environment and switching to "cautious mode" in a natural environment with a high cost of undesirable situations. Nevertheless, the variation algorithm is formed by the system designer. The behavior of systems that differ in the choice of compromise will be different, leading to distinct "trajectories" of accumulation of experience.

In the case of a material/continuous environment, actions have, in addition to the two ratings mentioned above, a third attribute - the duration of execution; states take on the tribute of the predicted moment of reaching the situation. The situation gets the attribute of the expected moment of reaching the situation; the criterion for situation evaluation should consider the value of such attribute.

The decision-making procedure described above uses *only the facts of the possibility of specific consequences*. In approaches that use probability estimation (see Probability in decision making), the situation becomes more complex. Situations are *attributed by the spectrum of possible ratings, each with its own estimate of the probability*. Accordingly, instead of single compromise parameter k, which determines the trade-off between utility and risk, a set of similar parameters should be used. This significantly complicates both the implementation of the calculations and assigning trade-off parameters by the system designer. Combined with the other aspects discussed in the above chapter, the usefulness of using this approach is highly subtle.

During the functioning of the AGI system, singular situations are possible and inevitable.

If for a specific situation, some of the actions are technically impossible or are prohibited for other reasons, they are simply excluded from the analysis.

The accumulated experience may not contain information about the possible consequences of some possible actions for any given situation. This is not an obstacle to applying the procedure described above, but it does provide an additional *opportunity to try out a previously untested action*. The implementation of such a decision *cannot be the*

result of formal analysis; an obvious option is a random choice between the action according to the described algorithm and one of the untested actions with the probability set by the system designer.

Situations are also possible when none of the allowed actions have been tested related to the current situation. In this case, the solution is to choose between inaction and a randomly chosen action. For a system that learns from scratch, the permanent choice of inactivity is obviously unacceptable; the decision is dictated by the system designer based on his own experience and intuition.

Finally, when there are several possible actions with identical evaluation criteria, the natural solution is random selection.

As you can see, the forecast and decision-making composition are closely intertwined with each other; technological aspects of implementation are non-trivial and will be described separately.

SUMMATION

- *Decision-making uses a multi-step forecast based on available experience.*

- *Possible actions include both atomic actions and sequences of atomic actions.*

- *The transition of the tested sequences of atomic actions into the category of possible actions expands the forecasting horizon.*

- *Each situation in the forecast is characterized by a preference/acceptability rating.*

- *Each possible action in the current situation is characterized by a pair of ratings (best and worst possible outcome) based on the forecast.*

- *Choosing a solution is a two-criterion optimization problem, which requires the choice of a compromise point by the system designer.*

- *Decision-making in a singular situation uses random choice.*

♡   💬4   ⤳

👤 Write a comment…

👤 **Brett Martensen** May 15   Liked by Mykola Rabchevskiy

Hi Mykola,

Very good article. Well written and easy to read.

Two small comments:

In your line that says: "situation -> action -> consequences -> new situation"

the consequences and new situation are really the same thing so

"situation -> action -> consequences = new situation"

might have been more accurate.

And your diagram of situations, actions and consequences could have used some action sequences

that converged on the same consequence because the situation to action is a one to many relationship, the action to consequence relationship is a one to many as you have on your diagram but also the consequence back to the action is a one to many relationship as well.

Cheers, Brett

♥ 1   Reply

> **3 replies by Mykola Rabchevskiy and others**

3 more comments…

Ready for more?

Subscribe