DISCLAIMER

This is a draft and is not the final product in neither substance or formatting.

Tracking the Truth in a Multi-Agent System by Belief Revision

Thomas Løye Skafte

January 31, 2021

Contents

1 Introduction

The concept of truth is widely accepted as the idea that there is some set of statements that are correct and a different set that are incorrect. The intention of collecting and monitoring these correct statements can then be said to be the intention of *truth tracking*. But what would be required of an agent to take up such a task?

One quality is that it must somehow be able to acquire information about the total set of statements while keeping its set of truths consistent. This concept is what is commonly known as *learning*. So it seems that if we want to create an agent that can track the truth it must be a *learning agent*.

In this body of work we will first introduce a way of creating such a learning agent initially introduced by Baltag, Gierasimczuk and Smets in [BGS19] and then go on to show a way of merging belief bases using a merging operator based on integrity constraints which was originally proposed by Konieczny and Pino Pérez in [KP02].

It will then be demonstrated how a set of multiple learning agents, all unable of tracking the truth individually, might be capable of the feat when combining their belief bases. Lastly we will show how a multi-agent system of universal learning agents do not lose their ability to identify in the limit any epistemic space when working together by means of a merging operator.

2 Learning Agents

In [BGS19] Beltag et al. shows a method of creating a *learning agent* that given an instance of their model it can learn everything that is learnable within that model. To understand how this is done we first need to grasp how the models are structured.

2.1 The Model

Considering that we want to work with belief revision, agents should have the capability to model not only the world they are in but also any world that is a possibility. For this purpose epistemic spaces are used.

Definition 1. Let $\mathbb{S} = (S, \mathcal{O})$ be an *epistemic space*, where S is a set of *possible worlds* and $\mathcal{O}$ is a set of *observable propositions*.

Consider $\mathcal{O}$ as the set of observations that the agent could encounter. Each instance of information, or observable proposition, is depicted as sets of all the worlds in S where the given information holds. Observation $O = \{s, t\}$ then holds for the worlds s and t. The set of all observations that are true in a given world s is noted as $\mathcal{O}_s = \{O \in \mathcal{O} \mid s \in O\}$. Thus $\mathcal{O}$ is a subset of the powerset of S, $\mathcal{O} \subseteq \mathcal{P}(S)$.

Belief revision is based on observations that the agents finds from the real world (sensors or other ways of gathering information). Each observation $O \in$

$\mathcal{O}$ is considered to be acquired one at a time and the sequence is known as $\sigma = (O_0, \ldots, O_n)$.

One can now find the worlds that are supported by each observation by $\bigcap \sigma$. If this set is of size one then we are in the trivial case where we can deduce the correct world. However if it is larger, then we are in a case of uncertainty and something more is required.

Epistemic spaces are a good way of modelling which worlds are possible, even considering the information the agent has gained, but when there are multiple candidates they do not provide any way to distinguish between more and less likely worlds. To incorporate belief some form of favouritism between the sets of worlds is required.

Definition 2. Let $\mathbb{B}_\mathbb{S} = (S, \mathcal{O}, \preceq)$ be a *plausibility space*, which is acquired by giving an epistemic space a *binary relation set* which is a total preorder $\preceq \subseteq S \times S$.

$s_0 \preceq s_1$ then means that s_0 is at least as likely as s_1 and from $\preceq$ being a total preorder it is also reflexive and transitive. The binary relation set can now be used for cases of uncertainty. Not only does it allow us to appoint a set of worlds as the most likely, it also lets us define a way of thinking of *belief*, as something that is true in all the most plausible worlds.

Definition 3. An agent believes p, Bp, when all states in the set of most likely worlds all satisfy p.

$$Bp \iff min_{\preceq} S \subseteq p$$
$$\text{where } min_{\preceq} X = \{t \in X \mid t \preceq s \text{ for all } s \in X\}$$

The way in which weight is assigned to the different worlds, meaning how they rank in the preorder, then decides what the agent believes in.

2.2 Belief Revision

In order to have agents aggregate towards a particular world, preferably the correct one, the model needs to be mutable.

Definition 4. Given a plausibility space $\mathbb{B}_\mathbb{S}$ and some observation $p \in \mathcal{O}$, $R_1(\mathbb{B}_\mathbb{S}, p) = \mathbb{B}'_\mathbb{S}$ is a *one-step revision method* that returns a new plausibility space.

Iterating over R_1 with an ordered list of observables σ gives a *belief revision method* $R(\mathbb{B}_\mathbb{S}, \sigma) = \mathbb{B}'_\mathbb{S}$.

$$R(\mathbb{B}_\mathbb{S}, \sigma * p) = R_1(R(\mathbb{B}_\mathbb{S}, \sigma), p)$$

σ is to be either an ordered infinite stream or finite sequence of observables. Allowing both options means that the two cases (1) having an up to date agent that is waiting for real world events to occur and (2) giving a backlog of observations to the agent, is handled in the same manner and is therefore indistinguishable. In the case where σ is empty the function simply returns the same $\mathbb{B}_\mathbb{S}$ as in the input.

2.2.1 Belief Revision Methods

Let us go over three revision methods introduced in [BGS19]. They were initially inspired by similar operations in DEL, which we will relate them to in case the reader knows of them already. The revisions methods R will be introduced by their one-step revision methods R_1.

The following plausibility space will be used to help explain the concepts by applying a single belief revision on the plausibility space and see how it changes it.

$S = \{s, t, r\}$
$\mathcal{O} = \{p, q\} = \{\{s, t\}, \{t, r\}\}$
$\preceq = \{t \preceq s, r \preceq t, r \preceq s\}$

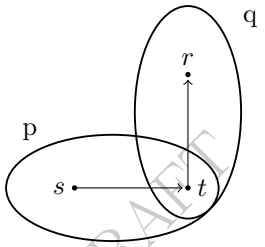

Figure 1: Initial plausibility space

The arrows indicate the binary relations of which world is deemed more likely with the world pointed at being preferred.

Conditioning

In DEL the *update* operation takes the new information as fact and removes any world that does not agree with the observation. This leads to shrinking the amount of possible worlds, removing uncertainty. Conditioning works the same way.

Definition 5. $Cond_1$ takes a plausibility space $\mathbb{B}_\mathbb{S} = (S, \mathcal{O}, \preceq)$ and an observation p and returns a new plausibility space $\mathbb{B}^p_\mathbb{S}$ where all worlds support p.

$$Cond_1(\mathbb{B}_\mathbb{S}, p) = (S^p, \mathcal{O}, \preceq^p)$$
$$\text{where } S^p = S \cap p, \quad \preceq^p = \preceq \cap (S^p \times S^p)$$

Applying the one-step revision operation $Cond_1$ on our example plausibility space together with the observation p returns a plausibility space without any of the worlds where p was not assigned to be true.

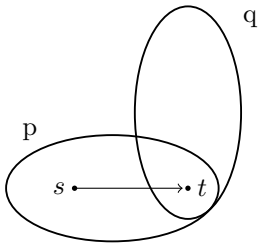

Figure 2: The plausibility space after applying $Cond_1(\mathbb{B}_\mathbb{S}, p)$

It should be pointed out that this method does not allow us to return to the previous state due to the nature of deleting. This is potentially useful for *hard information*, meaning observations that there can be no doubt of, and there is no point to even consider them.

Lexicographic Revision

First consider this method of combining two distinct binary relation rankings.

Definition 6. Consider $\preceq_i = (S^i \times S^i)$ and $\preceq_j = (S^j \times S^j)$ as two distinct binary relation sets, $S^i \cap S^j = \emptyset$. Let the following operation be a *concatenation* of two epistemic binary relation sets.

$$\preceq = [\preceq_i; \preceq_j]$$

Where in $\preceq$ every state of S^i is considered more entrenched than every state in S^j. The relations within $\preceq_i$ and $\preceq_j$ are kept as they were before the concatenation. This produces a total preorder $\preceq$.

Lexicographic revision, *radical upgrade* in DEL, refrains from deleting worlds and instead simply reorders the binary relation set $\preceq$ in favour of the new information, while keeping previous information whenever possible.

Definition 7. $Lex_1(\mathbb{B}_\mathbb{S}, p)$ elevates every world in S^p to be more entrenched than any world from $S^{\overline{p}}$ while keeping the binary relations between the worlds in S^p as well as the relations in $S^{\overline{p}}$.

$$Lex_1(\mathbb{B}_\mathbb{S}, p) = (S, \mathcal{O}, [\preceq^p; \preceq \setminus \preceq^p])$$

where $\preceq \setminus \preceq^p$ is the set subtraction containing every relation in $\preceq$ except the ones in $\preceq^p$.

This time we will see how the operation $Lex_1(\mathbb{B}_\mathbb{S}, p)$ changes the plausibility space to keep the ordering within p and reorder such that every world within p is considered more likely than r, which was previously the most believed world.

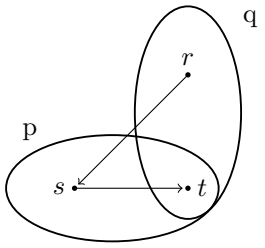

Figure 3: Example plausibility space after $Lex_1(\mathbb{B}_\mathbb{S}, p)$

Minimal Revision

Instead of elevating all the worlds that satisfy p as in Lex_1, *minimal revision* takes the most entrenched worlds in S^p and promote them to be the most entrenched worlds in all of S. In DEL this is known as *conservative upgrade*.

Definition 8. Let $\preceq^p_{min} = (min_{\preceq} S^p \times min_{\preceq} S^p)$ be the epistemic binary relation of the set of worlds that are the most entrenched in S^p.

$$Mini_1(\mathbb{B}_\mathbb{S}, p) = (S, \mathcal{O}, [\preceq^p_{min}; \preceq \setminus \preceq^p_{min}])$$

Applying the same observation to the initial example plausibility space as was done with Lex_1 we get a different result using $Mini_1$. Because t is the most likely world in p it is moved to be the most likely world overall and the $r \preceq s$ binary relation is kept.

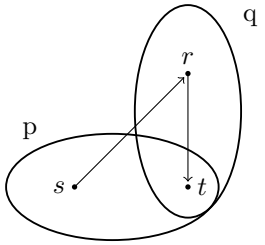

Figure 4: Example plausibility space after $Mini_1(\mathbb{B}_\mathbb{S}, p)$

2.3 Learning Method

It is obvious that the output of Lex_1 and $Mini_1$ depends heavily upon the entrenchment ordering of the plausibility space used and changing it would result in a different output. Extrapolating this to the iterated versions Lex and $Mini$ means that the initial state of the binary relations, before any belief revision is done, is also of high significance. Arguably this is even more so the case for $Cond$ since the initial ordering of the binary relations are never changed.

In Definition 4 belief revisions are defined to take a plausibility space as argument, and this is obviously necessary, but it means that an opinion on which worlds are more entrenched is required.

This begs the question of how this opinion should be formed. Considering an epistemic space $\mathbb{S} = (S, \mathcal{O})$ it is easy to argue where we get S and $\mathcal{O}$ from, but from where is the initial binary relation $\preceq$ in a plausibility space acquired? To help answer this some more definitions are necessary.

Definition 9. A mapping of the states in $\mathbb{S}$ to a binary relation set $\preceq$ is called a *prior plausibility assignment.*

$$\mathcal{P}(\mathbb{S}) = (S, \mathcal{O}, \preceq)$$

Definition 10. Let the usage of a belief revision method R and a prior plausibility assignment $\mathcal{P}$ on an epistemic space along with a sequence or stream of data σ be known as a *Learning Method*, L.

$$L_R^{\mathcal{P}}(\mathbb{S}, \sigma) = min\ R(\mathcal{P}(\mathbb{S}), \sigma)$$

Now the question of how the initial binary relations of $\preceq$ should be ordered can be explained by choosing a suitable prior plausibility assignment.

One of the main findings in [BGS19] is a method for selecting such a prior plausibility assignment based on the epistemic space $\mathbb{S}$ that it relates to.

Lets take a moment to consider what we are asking of our learning method. Specifically we want a learning method that can *identify within the limit* the correct world, no matter what epistemic space we throw at it. The concept of *identify within the limit* touches on the idea that we can only track the truth as well as the input allows us. So the learning method should *identify* the correct world *within the limits* of the input.

Definition 11. An epistemic space $\mathbb{S} = (S, \mathcal{O})$ is *identifiable in the limit by learning method* L if for every sound and complete data stream σ for each $s \in S$, $L(\mathbb{S}, \sigma)$ eventually outputs only $\{s\}$ henceforth.

Definition 12. Consider an epistemic space $\mathbb{S}$ that is identifiable within the limit by learning method L, then for identified world $s \in S$ there must exists at least one data sequence σ such that $L(\mathbb{S}, \sigma)$ only outputs $\{s\}$ from then on, no matter what new data that satisfies s occurs afterwards. Such a sequence σ is considered a *locking sequence.*

A locking sequence σ does not need to be the most direct route towards locking the output of the learning method. In fact, appending any data sequence τ, that is sound and complete w.r.t. s, to the end of σ resulting in the data sequence $\sigma * \tau$ is still a locking sequence for L and s due to the nature of locking sequences forcing it to still only return $\{s\}$.

A little bit surprising is how this relates to a different concept known as *tell-tale sets.*

Definition 13. A *finite tell-tale set* D_s is a finite set of observations, where $s \in \bigcap D_s$ and for any $t \in S$ where $t \in \bigcap D_s$, the set of observables required for s must be a subset of the set required by t, or else $t = s$.

Any other world $r \notin \bigcup D_s$ disagrees with s, and any other world in D_s, on one or more observation from D_s. We call these tell-tale sets because if they are assigned as true, then it is a hint that the state they belong to could be the correct one, but does not necessarily since it is possible there exists a world t that differs from s but is still within $\bigcap D_s$.

The two are related because if you take the set of a locking sequence σ for world $s \in S$, in order to remove duplicates, then that set will be a finite tell-tale D_s for the world s.

Definition 14. Let there exist a total map $D : S \to \mathcal{P}(\mathcal{O})$, such that $s \mapsto D_s$, where D_s is a finite tell-tale.

Using the previously mentioned method for converting locking sequences into finite tell-tales we get such a total map D.

The goal of generating this mapping is to create a partial ordering from the case when both $s, t \in D_s$, by preferring the one that requires the least assumptions, namely s. To handle this formally we introduce the *ordering tell-tale map* D'.

Definition 15. An *ordering tell-tale map* is a total map, $D' : S \to \mathcal{P}(\mathcal{O})$, such that $s \mapsto D'_s$, where D'_s is a finite tell-tale.

There is also a related injective map $i : S \to \mathbb{N}$, where if $t \in \bigcap D'_s$ and $\mathcal{O}_s \not\subseteq \mathcal{O}_t$, then $i(s) < i(t)$.

From this definition there seems to be little differentiating D and D', but the difference comes in having to uphold the the comparison requirement on the injective map. The way in which D' is created is by first finding D and then handling each specific case that goes against the injective map requirement by adding new information to D_s.

Consider $s, t \in D_s$, where $s \neq t$. i being an injective map means that it must induce an ordering. Take s and t such that $i(s) > i(t)$ and $\mathcal{O}_s \not\subseteq \mathcal{O}_t$, then we do not uphold the requirement. However, since $\mathcal{O}_s \not\subseteq \mathcal{O}_t$ we know that there is some proposition $p \in \mathcal{O}_s$ that is not in $\mathcal{O}_t$. If we add p to D_s then we uphold the requirement again, since they are not both in D_s. The new finite tell-tale that is the combination of D_s and p is D'_s.

The reasoning for needing D' is that if $s, t \in D_s$ and $\mathcal{O}_s \not\subseteq \mathcal{O}_t$, where s is the true world, then since any information gained from σ supports s, and therefore t as well, σ will never provide a way to differentiate between s and t. The solution in D' is akin to Occam's razor in that we choose the world that requires the least amount of assumptions as the more likely option. Using D' we can create a partial order by introducing the comparison method $\preceq^1_{D'}$ for $s, t \in S$,

$$\text{iff } t \in \bigcap D'_s \text{ then } s \preceq^1_{D'} t$$

From which we can get $\preceq_{D'}$ by transitive closure of the relation $\preceq^1_{D'}$. With use of the Order-Extension Principle [Man20] we can extend $\preceq_{D'}$ to a total order that in turn can function as the initial binary relation $\preceq$ for $\mathbb{B}_{\mathbb{S}}$.

Hereby we have a procedure that can be used as a prior plausibility assignment $\mathcal{P}$, given we know a locking sequence for each world.

Lemma 2.1. A binary relation $\preceq$ can be generated from the epistemic space $\mathbb{S}$ by applying the following prior plausibility assignment $\mathcal{P}$

1. Acquire a locking sequence σ_i for each world.
2. Calculate each worlds ordering tell-tale D'_i from $D_i = set(\sigma_i)$.
3. From D' find a partial ordering and extend it to a total preorder.

Definition 16. A learning method L is *universal* if it can identify within the limit any epistemic space that is identifiable within the limit.

Theorem 2.2. Using $\mathcal{P}$ together with either *Lex* or *Cond* produces a *universal learning method*.

Mini does decidedly not providing a *universal learning method* with the prior plausibility assignment $\mathcal{P}$.

A proof can be found in [BGS19].

3 Multiple Agent Belief Revision

The normal goal of multi-agent systems is to exploit the multiple perspective nature to either enhance the performance or even solve new problems with this advantage. The idea with MAS in belief revision is handling the situation of individual agents not having enough information alone, but combined they do.

One solution to this problem is simply forwarding the observations to a trusted agent that can apply singular agent belief revision, as per [BGS19], on the collective information. This centralised concept we will call *multiple source belief revision*, MSBR.

Since the model can only handle a single observation at a time, it is required that the observations are ordered in some manner. As long as the observations are non-erroneous the way in which the serialisation occurs does not matter, and while we do not require them to be complete, they must still be sound individually. The interesting part is that the combined observations must still provide a sound and complete data sequence with regard to the correct world in order to be identifiable within the limit. This is the same requirement we have for a normal single agent belief revision model.

Using MSBR we can handle multiple sources with the known model and learning methods that follow it. But there are still reasons to investigate other solutions. One limitation of MSBR is that it does not allow for any privacy between agents since each agent is forfeiting all observations made. Say an observation provides some side information besides the goal of the truth tracking,

that the agent would prefer to keep private, then the centralised agent would be just as capable of calculate that information from the observation.[LW01]

The objective behind combining the observations in a centralised agent is that the information that each agent has needs to be combined in some fashion, and the purest method of doing this is by combining the observations. However the currently most believed result of the learning methods output still contains the information obtained from the observations. So another possible point in the chain of truth tracking where information merging could occur is with the currently believed result. It turns out that this method has its own problems but we will address those later.

3.1 Belief Merge

In order to combine the information from the outputs of the learning agents an efficient method of aggregating multiple worlds into a singular collective answer is required. To the pursuit of this purpose the classical notion of *belief merge* from [Pig16] is introduced, allowing for merging of n agents presenting their belief as belief bases.

Definition 17. Let a *belief profile*, E, be a multi-set of belief bases.

$$E = \{K_1, \ldots, K_n\}$$

Then a *belief merger*, Δ_{IC}, is a aggregation function that takes a belief profile and outputs a new belief base K_c which satisfies a set of integrity constraints IC. K_c then represents what the collective believes to be true.

$$\Delta_{IC}(E) = K_c$$

The standard set of integrity constraints that are required for Δ_{IC} to be a belief merge operator, originally from [KP02], are the following.

(IC0)	$\Delta_{IC}(E) \models IC$
(IC1)	If IC is consistent, then $\Delta_{IC}(E)$ is consistent.
(IC2)	If $\bigwedge E$ is consistent with IC, then $\Delta_{IC}(E) \equiv \bigwedge E \wedge IC$
(IC3)	IF $E_1 \equiv E_2$, and $IC_1 \equiv IC_2$, then $\Delta_{IC_1}(E_1) \equiv \Delta_{IC_2}(E_2)$
(IC4)	If $K_1 \models IC$ and $K_2 \models IC$, then $\Delta_{IC}(\{K_1, K_2\}) \wedge K_1$ is consistent if and only if $\Delta_{IC}(\{K_1, K_2\}) \wedge K_2$ is consistent.
(IC5)	$\Delta_{IC}(E_1) \wedge \Delta_{IC}(E_2) \models \Delta_{IC}(E_1 \sqcup E_2)$
(IC6)	If $\Delta_{IC}(E_1) \wedge \Delta_{IC}(E_2)$ is consistent, then $\Delta_{IC}(E_1 \sqcup E_2) \models \Delta_{IC}(E_1) \wedge \Delta_{IC}(E_2)$
(IC7)	$\Delta_{IC_1}(E) \wedge IC_2 \models \Delta_{IC_1 \wedge IC_2}(E)$
(IC8)	If $\Delta_{IC_1}(E) \wedge IC_2$ is consistent, then $\Delta_{IC_1 \wedge IC_2}(E) \models \Delta_{IC_1}(E) \wedge IC_2$

(IC0) ensures that the result follows the integrity constraints. (IC1) if the constraints are consistent then so is the result. (IC2) states that when its possible the result of the merge is the conjunction of the profile and the integrity

constraints. (IC3) irrelevancy of syntax; if two belief bases are logically equivalent and two integrity constraints are equivalent, then the result of the two will be logically equivalent. (IC4) is the principle of fairness; when merging two or more belief bases the merger must not give preference towards any of them. (IC5) if two groups agree on some assignment of a proposition, then when joining the two the assignment must be chosen. From (IC5) and (IC6) together we gain that if it is possible to find two subgroups that agree on an assignment, then the result of the merge will include that assignment. (IC7) and (IC8) provide the notion of closeness is well-behaved; given that merging E under IC_1 and IC_2 is consistent, then IC_1 will remain fulfilled if more restrictions, such as IC_2 is added. This also means that if a merge results in K_c then adding more restrictions will not change the result. Intuitively this can be understood as; adding restrictions can only remove options and removing some option does not degrade the comparative value of K_c or increase the comparative value of a third option. For example, say there are three options that a merger Δ_{IC_1} could choose from, namely K_a, K_b and K_c. If $\Delta_{IC_1}(E) = K_c$, then re-running the merging with additional restrictions IC_2, that removes K_b as an option, will always result in $\Delta_{IC_1+IC_2}(E) = K_c$.

These were the standard postulates that are required, but alone they do not provide much of a merger. To gain anything of use we must add additional postulates. This is the design choice that comes with integrity constraint based mergers, and it gives a lot of freedom.

The merger that we will be looking into is the *majority merger*, for which the goal is that the chosen result must be the most supported one. This does not mean that the merger only picks between the belief bases in profile, but rather it picks the sound belief base that is closest to the profile as a whole.

Definition 18. If, for every integer n, E^n expresses the multi-set containing n times E, then a merging operator from the majority merger class Maj satisfies the following postulate (Maj) in addition to (IC0-8).

$$\text{(Maj)} \mid \exists n \Delta_{IC}(E_1 \sqcup E_2^n) \models \Delta_{IC}(E_2)$$

3.2 Distance Based Merging

With what we have discussed so far integrity constraint mergers are nothing more than an aggregation function with a set of postulates that are desirable, however it is not too difficult to implement a procedure that upholds them.

As mentioned earlier the majority merger selects the world that is the closest to the whole profile. The notion of distance imply that it is possible to measure some value that represents the difference between worlds, and that is exactly what is done.

There are all of two requirements for the chosen distance function, d. It must be symmetric and if the distance is zero the two worlds must be the same world. Otherwise any arbitrary distance function can be chosen, for example Hamming distance.

Distance Function, $d(w, w')$
$d : W \times W \to R^+$, s.t. for all $w, w' \in W$:

$$d(w, w') = d(w', w)$$
$$d(w, w') = 0 \text{ iff } w = w'$$

The distance function is then used in pursuit of applying a total ordering on the set of sound worlds, where the minimum world would then be the closest world to the belief profile. To do that we need a way of comparing a world w to a belief base K. Consider a belief base as set of observations that are believed to be true, then for any belief base there exists a set of worlds, $mod(K)$, that satisfy K. The distance between w and K is then the shortest distance between w and any world $w' \in mod(K)$. Formally this can be written as

$$d(w, K) = min_{w' \in mod(K)} d(w, w')$$
$$mod(K) = \{w \in W \mid w \models K\}$$

Being capable of calculating the distance between a world and a belief base lets us find the distance from a world w to each belief base K in the profile E, however it is not obvious how the distance from w to E should be determined. This is where the *collective distance function* $D(w, E)$ comes in, which is yet another modular part of the IC merging system.

Collective Distance Function, $D(w, E)$
Aggregation function, $D : R^{+n} \to R^+$

$$D(w, E) = D(d(w, K_1), d(w, K_2), \ldots, d(w, K_n))$$

The conceptual idea behind D is that it should provide a means to combine the distances w has to each belief base in E, such that we can compare it to other worlds $w' \in W$. Additionally the user is provided another opportunity to select the function to provide the functionality that is sought. Simple example is to sum each distance to a total distance. The only requirements there are for D are

1. If $x \geq y$, then $D(x_1, \ldots, x, \ldots, x_n) \geq D(x_1, \ldots, y, \ldots, x_n)$
2. $D(x_1, \ldots, x, \ldots, x_n) = 0$ if and only if $x_1 = \cdots = x_n = 0$
3. $D(x) = x$

$D(w, E)$ lets us compare worlds by their distance to E which in turn provides a method to acquire a total pre-order by calculating $D(w, E)$ for all $w \in W$.

The result of the merging is then found by selecting the world with the smallest distance to the profile, which is the same as selecting the smallest world in the ordering.

3.2.1 Minisum Belief Merger

The *minisum* merger $\Delta_{\sum}$ is an instance of such a integrity constraint belief merger that is identified by a specific distance function d and collective distance function D.

$$d(w, w') = \text{Ham}(w, w')$$
$$D(w, E) = D(d_1, \ldots, d_n) = \sum_i d_i$$

Hamming Distance, Ham(w, w')
The hamming distance between two worlds w and w' is calculated by the amount of propositions the two disagree on.

$$\text{Ham}(w, w') = |prop(w) \setminus (prop(w) \cap prop(w'))|$$

Minisum satisfies both the standard integrity constraint postulates IC0-8 as well as the majority postulate Maj, and as such is considered a *majority merger*.

3.2.2 Leximax Belief Merger

$\Delta_{leximax}$ also uses the hamming distance for d, but instead of doing a summation for D it orders the distances in reverse lexicographical order, also known as descending order.

$$d(w, w') = \text{Ham}(w, w')$$
$$D(w, E) = D(d_1, \ldots, d_n) = lex(d_1, \ldots, d_n)$$

The world that $\Delta_{leximax}$ then crowns the result is the world with the smallest distance in the first index in the *lex* result. What this equates to is selecting the world w_c that minimising the distance between w_c and the belief base $K \in E$ that is the furthest from w_c.

This is known as a *arbitration merger* and it differs from the majority merger by selecting the result on a different basis. While the majority merger selects the most supported world, an arbitration merger selects the world that minimises the incorrectness of the least correct belief base. This can also be thought of as minimising the dissatisfaction of the most dissatisfied agent.

$$\left.\begin{array}{l} \Delta_{IC_1}(K_1) \leftrightarrow \Delta_{IC_2}(K_2) \\ \Delta_{IC_1 \leftrightarrow \neg IC_2}(K_1 \sqcup K_2) \leftrightarrow (K_1 \leftrightarrow \neg K_2) \\ IC_1 \not\models IC_2 \\ IC_2 \not\models IC_1 \end{array}\right\} = \Delta_{IC_1 \wedge IC_2}(K_1 \sqcup K_2) \leftrightarrow \Delta_{IC_1}(K_1) \tag{Arb}$$

When we introduced the D function it was said to be an aggregation function, $D : R^{+n} \to R^+$, but the D that $\Delta_{leximax}$ uses does not strictly follow this. It still works because we also change how the total order is devised. This shows the flexibility and modular nature of the scheme.

4 Merging Learning Methods

The goal of introducing the integrity constraint merging concept is to have a way of combining the results that different learning agents output. The idea being that the information that the individual agents acquire from their data sequences persists through the learning methods. The intention is then to combine the information that is present in the learning agent's output and thereby making a multi-agent system that can identify the true world in more situations than the agents would individually.

For ease of reference we will be addressing single agent belief revision as SBR and multi-agent belief revision as MABR.

By definition 11 the learning agents are said to have identified a world in the limit when it outputs that world only continuously onwards. So if the world does output a singleton then that agent is considered to have solved the truth tracking. The situation we are interested in is when the agents cannot do it alone. Therefore we will from here on be addressing the situation in which the output of the agents are multiple worlds unless otherwise explicitly noted.

The integrity mergers that we will be applying take a belief profile as input, however the learning methods output a set of worlds. Each of these worlds are considered of equal probability to be the correct one by the agent and they all contain the information gained by the observations of said agent. Conceptually, the differences between these worlds represent the information that the agent is uncertain of and the intersection of the worlds is the information that the agent believe to be correct from the observations along with the prior plausibility assignment. The intersection can also be thought of as the belief base of the agent, at that specific point when the output is given.

Consider now the situation of trying to track the truth in $\mathbb{S}$ with n agents each with their own data sequence σ_i. Let $L_R^{\mathfrak{P}}(\mathbb{S}, \sigma_i)$ be the set of worlds that agent i outputs from its learning method, meaning $K_i = \bigcap L_R^{\mathfrak{P}}(\mathbb{S}, \sigma_i)$ is the belief base of agent i.

Definition 19. A world $s \in S$ is *identified in the limit by merger* m if m always selects s as the output for all belief profiles $E = \{K_1, \ldots, K_n\}$, for which it holds that

1. $\forall K_i \in E$, K_i is based on SBR on data sequence σ_i from agent i
2. σ_i is sound w.r.t. s
3. The concatenation of all data sequences, $\sigma_A = \sigma_i * \cdots * \sigma_n$, is complete w.r.t. s.

$$\Delta_{IC}(K_1, \ldots, K_n) = s$$

Definition 20. A merger m is a *universal merger on class C of epistemic spaces* if it can identify in the limit every epistemic space s in C. A *universal merger* is universal on the class of all epistemic spaces.

Unless otherwise noted we will be referencing to the class the majority mergers when talking about mergers from here and onwards.

Let us examine the case where each σ_i is sound, but not complete, w.r.t. s. Now every agent is incapable of identifying the true world by itself, because some crucial information is missing. In the same vein, if that essential observation is missing from every agents σ_i, then the merger has no hope of ever acquiring that information. This is obvious considering the nature of a merger is to combine information and not acquiring new information.

Lemma 4.1. In order for $s \in S$ to be identified within the limit by majority merger *Maj* from the profile $E = \{K_1, \ldots, K_n\}$ the concatenation of the observations $\sigma_A = \sigma_1 * \cdots * \sigma_n$ must be sound and complete w.r.t. s.

Proof. Say s is identified within the limit by merger *Maj* on E, but σ_A is either not sound or not complete w.r.t. s.

σ_A not being sound w.r.t s means some observation O goes against s, and nothing is stopping this to occur for every agent. This leads to a majority not supporting O, which is required for s so the merger will pick some other world. The possibility of not picking s means we do not have a guarantee of picking s and s is therefore not identified in the limit, contradiction.

σ_A not being complete w.r.t. s means there is some observation in $\mathcal{O}_s$ that is not found in any of the data sequences σ_i. Each learning agents decision on this observation is decided by the prior plausibility assignment, which could go against the truth assignment in s. No guarantee. □

Lemma 4.2. For s to be identified within the limit by majority merger *Maj* from the profile $E = \{K_1, \ldots, K_n\}$, s must also be identified in the limit by a SBR agent using a universal learning method on σ_A.

$$\Delta_{Maj}(K_1, \ldots, K_n) = s$$
$$\Longleftrightarrow$$
$$L_R^{\mathcal{P}}(\mathbb{S}, \sigma_A) = \{s\}$$

Proof.
$\Longrightarrow$: From Lemma 4.1 we know that σ_A must be sound and complete w.r.t. s for s to be identifiable within the limit by majority merger *Maj* and from theorem 2.2 we know that if the data sequence is sound and complete w.r.t. s then s is identifiable within the limit by a universal learning method $L_R^{\mathcal{P}}$.
$\Longleftarrow$: For the reverse proof we use the same logic, but in the opposite direction. □

The implications of this is that any truth tracking done with a majority merger can be performed by a SBR agent, if that agent is provided with the same amount of information as all of the agents combined.

Theorem 4.3. Any majority merger is a universal merger.

Proof. From lemma 4.2 we know that any epistemic space identifiable in the limit by a SBR agent, can also be identified by a MABR setup. Since the SBR learning agent is universal and can identify every world identifiable in the limit, then the majority merger is also capable of identifying every epistemic space that is identifiable in the limit. □

References

[BGS19] Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. "Truth-Tracking by Belief Revision". English. In: *Studia Logica: An International Journal for Symbolic Logic* 107.5 (2019), pages 917–947. ISSN: 0039-3215. DOI: 10.1007/s11225-018-9812-x.

[KP02] Sébastien Konieczny and Ramón Pino Pérez. "Merging Information Under Constraints: A Logical Framework". In: *Journal of Logic and Computation* 12.5 (October 2002), pages 773–808. ISSN: 0955-792X. DOI: 10.1093/logcom/12.5.773. eprint: https://academic.oup.com/logcom/article-pdf/12/5/773/3852854/120773.pdf. URL: https://doi.org/10.1093/logcom/12.5.773.

[LW01] Wei Liu and Mary Anne Williams. "A framework for multi-agent belief revision". eng. In: *Studia Logica* 67.2 (2001), pages 291–312. ISSN: 15728730, 00393215. DOI: 10.1023/A:1010555305483.

[Man20] Michael Mandler. "A Quick Proof of the Order-Extension Principle". In: *The American Mathematical Monthly* 127.9 (2020), pages 835–835. DOI: 10.1080/00029890.2020.1801081. eprint: https://doi.org/10.1080/00029890.2020.1801081. URL: https://doi.org/10.1080/00029890.2020.1801081.

[Pig16] Gabriella Pigozzi. "Belief Merging and Judgment Aggregation". In: *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University, 2016.