Autobiography Based Prediction in a Situated AGI Agent

Ladislau Bölöni

Dept. of Electrical Engineering and Computer Science
University of Central Florida
4000 Central Florida Blvd, Orlando FL 32816
lboloni@eecs.ucf.edu

Abstract. The ability to predict the unfolding of future events is an important feature of any situated AGI system. The most widely used approach is to create a *model of the world*, initialize it with the desired start state and use it to *simulate* possible future scenarios. In this paper we propose an alternative approach where there is no explicit model building involved. The agent memorizes its personal autobiography in an unprocessed narrative form. When a prediction is needed, the agent *aligns* story-lines from the autobiography with the current story, *extends* them into the future, then *interprets* them in the terms of the current events. We describe the implementation of this approach in the Xapagy cognitive architecture and present some experiments illustrating its operation.

Keywords: Situated agent, Prediction, Narratives.

1 Introduction

The ability to reason about the future (to make predictions in real or hypothetical situations) is admittedly one of the key components of any situated AGI system. A widely used way to perform such predictions is through *model building* coupled with *simulation* (this corresponds to claim 6 made for the CogPrime design of future AGI [5]). We create a model describing how the world operates. Whenever we want to identify whether a certain plan would succeed, or get a likely series of events from a starting point, we bring the model to the initial conditions, and allow it to simulate the unfolding events. The predictions can then be read out from the results of the simulation. For a situated agent which must continuously predict the future state of the world, the approach will follow the following algorithm:

```
Offline:
   MODEL Build a model out of data (and a priori knowledge)
Online:
   Repeat:
      Sense the state of the environment
      INITIALIZE the model with the current state
```

B. Goertzel et al. (Eds.): AGI 2014, LNAI 8598, pp. 11–20, 2014.

```
    SIMULATE by running the model
    READ-OUT the state of the model as a prediction
    [optional] Update the model based on new recordings
```

Thus, the approach consist of the offline MODEL step and the INITIALIZE-SIMULATE-READOUT online cycle. In some cases we run multiple models in parallel with different assumptions (e.g. uncertain sensing). Often, but not always, the model is *learned* based on training data. Note that the online model update step is optional – in fact, in many applications it is considered undesirable as it introduces unpredictability in the future behavior of the agent.

In this paper we describe a radically different approach to prediction. We build no model and there is no offline or online learning involved. The unprocessed data sensed by the agent is recorded as *stories* in the *autobiographical memory* (AM). The prediction cycle will look as follows:

```
Offline:
   << nothing >>
Online:
   Repeat:
      Sense the state of the environment
      ALIGN stories from the AM with the current state
      EXTEND the aligned stories into the future
      INTERPRET the extended stories in terms of the current state
      [optional] Record the current events in the AM
```

The online recording of the current events is optional, just like the online learning for the model-based prediction.

1.1 Predictive Power and Performance: Does This Even Make Sense?

The proposed AM-based prediction immediately raises a number of questions. Can it match the predictive power of the model-based approach? Isn't the model-based approach vastly more efficient? Does this make any sense?

Let us discuss first the theoretical limits of the predictive power. The sources of the model can be (a) scientific and engineering knowledge and (b) experimental data. Both of these can be expressed in narrative form: humans learn science and engineering from books and lectures and the setup and results of experiments can also be described as stories. Thus, the model-based and the AM-based approach can operate on the same source of information: the first compiling it into a model, while the second merely storing it in a narrative form. If we really, desperately want to match the model-based approach, we can (a) assume that all stories are relevant in the align step and (b) hide a just-in-time model building algorithm in the interpretation step. Naturally, emulating the model-based approach this way is highly inefficient, as the model is built not once per agent, but once per time-step.

The question about the performance of the AM-based approach boils down to (a) whether we can afford to carry and store the full AM and (b) how many stories are relevant at any given moment?

If the source of information is “Big Data”, such as “all the data humanity had ever produced”, this obviously creates major problems for the ALIGN step. If, however, our ambitions are limited to matching human intelligence, we need a much smaller AM. A human does not operate on all the data ever produced by humanity, but only on his/her personal experience, and this can be of a very moderate size. If we write up a narrative from a human life experience, at the rate of 1 sentence/second, we end up with 600 million sentences for a 30 year old person, a large but manageable number.

If we consider how many stories are relevant in a given circumstance, the number is much smaller. For instance, an airline pilot is required to have 1500 flight hours, and the experience of a trial lawyer can be counted in at most hundreds of cases. Naturally, this first person experience is complemented by the books read by the pilot or the lawyer.

Still, wouldn’t the extracted models be a more compact and elegant representation? In certain areas, certainly. Many of us suffer from “physics envy” and hope to discover beautiful, compact formalisms similar to Newton’s laws or quantum mechanics which capture vast domains of reality in several equations. Turns out, however, that few fields have such compact models. For instance, there is reason to believe that a general model of human behavior as an individual and as a social agent would have a state space larger than that of the personal experience of a single human. This might explain why the field of sociology often proposed [10] but never succeeded [8] in building a general model of social behavior.

1.2 A Running Example and Model Based Solutions

Let us now consider a simple situation which we will use as a running example:

> *Robby the Robot is currently watching on the TV a dramatization of Homer’s Iliad. On the screen he sees the fight between Hector and Achilles, while the voice-over narration comments on the story. Robby fears that the story will end in the death of one of the characters. Suddenly, the program is interrupted by a commercial. Frustrated, Robby tries to envision a way in which the story will end peacefully.*

While “pure logic” will not help Robby in this scenario, both the model based approach and the autobiography based approach would be able to generate Robby’s behavior, albeit in very different ways.

A model based approach would need a model of the one-to-one combat of the type in which Hector and Achilles is engaged on. There are several ways to implement this. In ACT-R[2,1] or Soar [7], this model would be represented using productions. Another possible approach is to use scripts to model the various possible scenarios [12]. Another approach would be to use variations of first order predicate logic, such as situation calculus, event calculus [9] or episodic logic [13] which allows the translation of the English language stories into a rich logical model. Finally, it is possible to develop probabilistic models for prediction, often in the form of conditional random fields (CRFs) or factor graphs as in the Sigma cognitive architecture [11]. There are also approaches which take the story as a primary component of the design of the system [15,4]. Nevertheless, in most of these systems, the interpretation of stories is done using a model representation.

In contrast to these approaches, the autobiography based approach does not need previous model building or learning. What it requires, however, is relevant autobiographical experience. In order to behave in the way described above, Robby must have had previous experience watching or reading about one-to-one combat. Furthermore, its personal experience will affect its predictions. If all the fights remembered by Robby had ended peacefully, the robot will not predict the death of a character. On the other hand, unless it had seen fights ending without the death of the looser, Robby will have difficulty outlining a way the fight can end without violence.

In the remainder of this paper we describe the ways in which the AM-based prediction is implemented in the Xapagy architecture and run some experiments.

2 Implementation

2.1 The Xapagy Cognitive Architecture

Xapagy is a cognitive architecture developed with the goal of mimicking the ways humans reason about stories. Stories are described in Xapi, a language that approximates closely the internal representational structures of the architecture but uses an English vocabulary. Xapi should be readable for an English language reader with minimal familiarity of the internal structures of Xapagy.

Xapi sentences can be in subject-verb-object, subject-verb or subject-verb-adjective form. A single more complex sentence exists, in the form of subject-communication verb-scene-quote, where the quote is an arbitrary sentence that is evaluated in a different *scene*. Subjects and objects are represented as *instances* and can acquire various attributes in form of *concepts*. Xapi sentences are mapped to objects called *verb instances* (VIs).

One of the unexpected features of Xapagy instances is that an entity in colloquial speech is often represented with more than one instance. These instances are often connected with *identity relations* but participate independently in VIs, shadows and headless shadows. We refer the reader to the technical report [3] for a “cookbook” of translating English paragraphs of medium complexity into Xapi.

The newly created VIs of a story are first entered into the *focus*, where they stay a time dependent on their salience, type and circumstances. For instance, VIs representing a relation will stay as long as the relation holds. On the other hand, VIs representing actions are pushed out by their successors. During their stay in the focus, VIs acquire salience in the *autobiographical memory* AM and are connected by *links* to other VIs present in the focus. After they leave the focus, VIs and instances cannot change, cannot acquire new links, and cannot be brought back into the focus.

2.2 The ALIGN Step: Shadowing

The technique of aligning story lines with the ongoing story in Xapagy is called *shadowing*. Each instance and VI in the focus has an attached *shadow* consisting of a weighted set of instances, and respectively VIs from the AM. The maintenance of the shadows is done by a number of dynamic processes called *diffusion activities* (DAs). Some of the DAs create or strengthen shadows based on direct or indirect attribute matching. For instance, Achilles will be matched, in decreasing degrees, by his own previous instances, other Greek warriors, other participants in one-to-one combat, other humans and finally, other living beings. More complex DAs, such as the scene sharpening and the story consistency sharpening DAs, rearrange the weights between the shadows. If a specific storyline is a strong match to the current one, the individual components will be matched as well even if their attributes are very different. The different DAs interact with each other: a shadow created by a DA can be strengthened or weakened by other DAs.

Very weak shadows are periodically garbage collected. To avoid filling the shadows with a multitude of weak shadows (which can happen in the case of highly repetitive but low salience events) the DAs use probability-proportional-to-size sampling without replacement [6,14] when bringing components of the AM into the shadow.

2.3 The EXTEND Step: Link Following

The AM of the agent consists of the VIs connected using links. The link types used in Xapagy are *succession*, *coincidence*, *context* (which connect a VI to the relations which held during their stay in the focus) and *summarization*. The extension of the shadows (matched and aligned stories) into the future is based on a triplet called the Focus-Shadow-Link (FSL) object. An FSL is formed by a VI in the focus F, a VI in its shadow S and a VI L linked to S through a link of a specific type. For instance, a succession-type FSL which appears in our story representation is:

```
F: "Achilles" / wa_v_sword_penetrate / "Hector".
S: "Mordred" / wa_v_sword_penetrate / "Arthur".
L: "Arthur" / changes / dead.
```

Normally, the agent generates up to several thousand FSL objects, each with their specific weight. The weight is a monotonic function of (a) the strength of F in the focus, (b) the shadow energy of S, (c) the strength of the link connecting S to L. Just like the shadows, the FSLs are maintained by DAs and vary in time.

2.4 The INTERPRET Step: Headless Shadows

The L component of the FSL will be our source for prediction. Intuitively, these components are VIs which happened in the story lines which shadow the current VIs, thus it is likely that something like this will happen this time as well. The problem, however, is that the L VI refers to the shadowing story line, not to the current scene. For instance, in our example L happens in the world of the Arthurian legend and it does not tell us anything about Hector and Achilles. We can infer that the FSL predicts the death of one of the combatants, but which one?

The solution is found by calculating the *reverse shadow* of the Arthur instance. While a (direct) shadow answers the question which AM instances, with what weight are aligned with a given focus instance, the reverse shadow determines for a given AM instance, which focus objects it shadows.

In our case we have:

```
ReverseShadow("Arthur") =
   0.11 "Hector"
   0.03 "Achilles"
```

We *interpret* the FSL by creating all the feasible combinations of interpretations of it, and weighting them according to the ratios in the inverse shadow. In our case the FLS will be exploded into two FSL Interpretation (FSLI) objects:

```
FSLI: I: "Hector"/changes/dead. w = 0.05 * 0.11 / (0.03+0.11)
FSLI: I: "Achilles"/changes/dead. w = 0.05 * 0.03 / (0.03+0.11)
```

We mentioned that the agent might maintain thousands of FSL objects, which might give raise to tens of thousands of FSLI objects. The number of predictions, however, is much smaller, because many FSLI objects will have the same or similar interpretation components. To capture this, we perform a similarity clustering over the FSLI objects, based on the interpretation component. This creates a smaller pool of possible interpretations, to which each of these FSLI objects act as a support. The overall shape of such a cluster is very similar to that of a shadow, with the exception that the head of the shadow (such as the Hector/changes/dead) event is *not yet instantiated* as a VI. We call this cluster a *continuation-type headless shadow* (HLS).

One of the challenging aspects of reasoning with HLSs is how to combine their supports, in particular how FSLIs with different link types strengthen or weaken the case for HSL. The simplest case in which we are making predictions about VIs expected in the future, succession, context and summarization links provide positive evidence. In contrast, the existence of the predicted VI in the focus and predecessor links (the inverse of successor links) provide negative evidence.

Although it is beyond the scope of this paper, we note that the same mechanism might be used for other reasoning processes beyond predicting future events. For instance, we can infer actions which happened in the past, but were missed by the sensing, relations which hold but had not been perceived and ways to summarize ongoing events.

3 Experiments

Our experiments involve a Xapagy agent which impersonates Robby from the scenario described in the introduction. We used the current version of the Xapagy architecture (1.0.366). To allow it to represent stories inspired from the Iliad, the agent was initialized with a collection of domain descriptions containing lists of concepts and verbs, as well as overlap and negation relationships between them. The domain description, however, does not attach any semantics to the verbs and concepts: the semantics must be acquired from the autobiography. We started with the core domains covering things such as basic spatial relations, naive physics, basic facts about humans and so on. For this set of experiments, we also created a specific domain ONE_TO_ONE_COMBAT listing concepts and verbs used in stories such as sword fight, sport fencing and boxing.

After initializing it with the domain, the agent was provided with a synthetic autobiography. This autobiography, beyond the generic part shared with other agents, included a set of set of stories specifically created for these experiments, providing the background for Robby's reasoning about the Achilles-Hector fight. This part of the autobiography included the fight between Hector and Patrocles, the fight when Achilles killed the Amazon Pentesilea and the fight when Hercules defeated but not killed the Amazon Hyppolyta. These battle-fights were complemented by the fight between King Arthur and Mordred at the battle at Camlann, where both were killed (according to one version of the legend). In addition, the autobiography included two generic fencing bouts ending with the weaker fencer conceding defeat, and the fencers shaking hand at the end of the bout. Finally, we included the box matches Cassius Clay vs Sonny Liston (1965) and Muhammad Ali vs George Foreman (1974).

3.1 The Duel of Achilles and Hector

Let us now see a representation of the main steps in the story seen by Robby on the television. The processing starts at timepoint t=8210 in the lifecycle of the agent.

```
8210  $NewSceneOnly #Reality,none,"Achilles" greek w_c_warrior,
          "Hector" trojan w_c_warrior
8211  "Achilles" / hates / "Hector".
8212  "Achilles" / wa_v_sword_attack / "Hector".
8213  "Hector" / wa_v_sword_defend / "Achilles".
8214  "Achilles" / wa_v_sword_attack / "Hector".
8215  "Hector" / wa_v_sword_defend / "Achilles".
8216  "Hector" / wcr_vr_tired / "Hector". // Marks Hector as tired
8217  "Achilles" / wa_v_sword_attack / "Hector".
8218  "Hector" / wa_v_sword_defend / "Achilles".
8219  "Achilles" / wa_v_sword_attack / "Hector".
8220  "Achilles" / wa_v_sword_penetrate / "Hector".
8221  "Achilles" / thus wcr_vr_victorious_over / "Hector".
8222  "Hector" / thus changes / dead.
```

While processing this story, the agent maintains its constantly evolving collection of shadows. To illustrate the operation of the shadow maintenance DAs, let us take a look at the shadows of Hector at the end of the story (t=8222), together with the shadow energy metric:

```
Shadows of "Hector" (end of scene with Achilles)
----------------------------------------
914.89 "Pentesilea" (scene with Achilles)
 32.63 weak fencer
 20.04 "Arthur" (scene with Mordred)
 14.28 strong fencer
  5.15 "Hector" (scene with Patrocles)
  4.82 Patrocles (scene with Hector)
```

To understand what the shadows signify, recall that in Xapagy entities which in colloquial speech are the same might be represented by different instances. Thus, the instance of Hector who killed Patrocles is not the same as the one who is fighting with Achilles (although they might be connected with an identity relation). This allows us to represent plans, fantasies, and alternative narratives - for instance, we can seamlessly represent the instances of King Arthur who was killed by Mordred at Camlann, the one who was mortally wounded and died at Camelot and the one who journeyed to the Isle of Avalon and is getting ready to return – which are all versions of the story. These instances will appear separately in the shadows. Usually, previous instances of the same entity will have a strong role in the shadow due to the similarities between the entities. What is surprising here is that the strongest shadow is not a previous instance of Hector, but that of Penthesilea. This illustrates the fact that the role played by the instance in the structure of the story (in this case: being on the loosing end of a fight with Achilles) matters more than the attributes (name, gender, nationality).

Let us assume that the television cuts to commercials at t=8219. At this moment, we have seen Hector becoming tired and Achilles launching an attack. The eight strongest continuation HLSs are:

```
0.964 Achilles / wr_vr_victorious_over / Hector.
0.482 Hector / changes / dead.
0.412 Hector / wa_v_concedes_defeat / Achilles.
0.389 Achilles / wa_v_sword_penetrate / Hector.
0.242 Achilles / wa_v_shakes_hand / Hector.
0.120 Hector / wa_v_sword_attack / Achilles.
0.052 Hector / wa_v_sword_penetrate / Achilles.
0.034 Achilles / wa_v_concedes_defeat / Hector.
```

The strongest prediction is that of the victory of Achilles while the second is that of the death of Hector. The list also contains some alternative scenarios, both of a peaceful termination, as well as that of a victory by Hector, albeit with a much weaker support.

If the agent would now try to imagine how the story unfolds, it would only need to instantiate internally the strongest continuation HLS. This would, of course, alter the shadows, and create a new set of continuation HLSs. By successively instantiating the strongest HLSs, we would obtain the following prediction:

```
8220  "Achilles" / wcr_vr_victorious_over / "Hector".
8221  "Hector" / changes / dead.
```

Which roughly corresponds to the way the story will unfold after the commercial break, albeit lacks details about the manner of Achilles killing Hector. In order to match the desired behavior where Robby tries to find a non-violent end, it can proceed by choosing to instantiate continuations which are typical to fencing bouts with friendly endings. In the following we list three timesteps, for each timestep showing the three strongest HLSs with the one chosen for instantiation marked with ***.

```
------ strongest continuations at t=8220.0 -----
0.964 "Achilles" / wcr_vr_victorious_over / "Hector".
0.482 "Hector" / changes / dead.
*** 0.412 "Hector" / wa_v_concedes-defeat / "Achilles".
------ strongest continuations at t=8221.0 -------
1.399 "Achilles" / wcr_vr_victorious_over / "Hector".
*** 0.505 "Achilles" / wa_v_shakes_hand / "Hector".
0.414 "Hector" / changes / dead.
------ strongest continuations at  t=8222.0 ------
*** 0.726 "Achilles" / wcr_vr_victorious_over / "Hector".
0.322 "Hector" / changes / dead.
0.159 "Achilles" / wa_v_sword-penetrate / "Hector".
--------------------------------------------------
```

So the overall prediction is now:

```
8220  "Hector" / wa_v_concedes-defeat / "Achilles".
8221  "Achilles" / wa_v_shakes_hand / "Hector".
8222  "Achilles" / wcr_vr_victorious_over / "Hector".
```

Notice that the continuation mechanism tries to maintain at least partial internal consistency. While we had to choose the third strongest HLS in the first timestep, we could choose the second one in the next one and the strongest one in the third one. At the same time, the HLS corresponding to the death of Hector is steadily diminishing at each step taken towards a peaceful turn of events.

4 Conclusions

In this paper we argue that for situated AGI agents, a prediction approach based on autobiography can be a complement or alternative to model-and-simulation based approaches. In particular, if the source of the agent's knowledge is exclusively his autobiographical experience, this approach can be both easier to build and more efficient than model based approaches. We have outlined how such an approach would work and experimentally demonstrated it using the Xapagy cognitive architecture.

References

1. Anderson, J., Bothell, D., Byrne, M., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111(4), 1036 (2004)
2. Anderson, J., Lebiere, C.: The atomic components of thought. Lawrence Erlbaum (1998)
3. Bölöni, L.: A cookbook of translating English to Xapi. arXiv:1304.0715 (2013)
4. Forbus, K., Riesbeck, C., Birnbaum, L., Livingston, K., Sharma, A., Ureel, L.: Integrating natural language, knowledge representation and reasoning, and analogical processing to learn by reading. In: Proc. of AAAI, pp. 1542–1547 (2007)
5. Goertzel, B., Ke, S., Lian, R., O'Neill, J., Sadeghi, K., Wang, D., Watkins, O., Yu, G.: The CogPrime architecture for embodied artificial general intelligence. In: IEEE Symposium on Computational Intelligence for Human-like Intelligence (CIHLI 2013), pp. 60–67. IEEE (2013)
6. Hanurav, T.: Optimum utilization of auxiliary information: π ps sampling of two units from a stratum. Journal of the Royal Statistical Society. Series B (Methodological), 374–391 (1967)
7. Lehman, J., Laird, J., Rosenbloom, P., et al.: A gentle introduction to Soar, an architecture for human cognition, vol. 4, pp. 211–253. MIT Press (1998)
8. Mills, C.W.: The sociological imagination. Oxford University Press (1959)
9. Mueller, E.: Understanding script-based stories using commonsense reasoning. Cognitive Systems Research 5(4), 307–340 (2004)
10. Parsons, T.: The social system. Psychology Press (1951)
11. Rosenbloom, P.S.: Rethinking cognitive architecture via graphical models. Cognitive Systems Research 12(2), 198–209 (2011)
12. Schank, R., Abelson, R., et al.: Scripts, plans, goals and understanding: An inquiry into human knowledge structures, vol. 2. Lawrence Erlbaum Associates Nueva Jersey (1977)
13. Schubert, L., Hwang, C.: Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language, 111–174 (2000)
14. Vijayan, K.: An exact π ps sampling scheme-generalization of a method of Hanurav. Journal of the Royal Statistical Society. Series B (Methodological), 556–566 (1968)
15. Winston, P.H.: The strong story hypothesis and the directed perception hypothesis. In: Langley, P. (ed.) Technical Report FS-11-01, Papers from the AAAI Fall Symposium, pp. 345–352. AAAI Press, Menlo Park (2011)