

Autobiography based prediction in a situated AGI agent

Ladislau Bölöni

Dept. of Electrical Engineering and Computer Science
University of Central Florida- Orlando, FL

August 1, 2014

1 Introduction

1 Introduction

2 Implementation

1 Introduction

2 Implementation

3 Experiments

1 Introduction

2 Implementation

3 Experiments

4 Conclusions

1 Introduction

2 Implementation

3 Experiments

4 Conclusions

Reasoning about the future

- Making predictions in real or hypothetical situations is an important component in any AGI system.
- The most widely used approach for prediction is **model building** followed by **simulation**

Claim 10 (g) Simulation is a good way to handle episodic knowledge (remembered and imagined). Running an internal world simulation engine is an effective way to handle simulation.
Ben Goertzel - CogPrime: An Integrative Architecture for Embodied Artificial General Intelligence

INITIALIZE-SIMULATE-READOUT

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

SIMULATE by running the model

READ-OUT the state of the model as a prediction

[optional] Update the model based on new recordings

A different model

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).

ALIGN-EXTEND-INTERPRET

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

Does this even make sense?

- Can it match the predictive power of the model-based approach?
- Isn't the model based approach vastly more efficient?

Theoretical limits of the predictive power

- Where do models come from?
 - ▶ scientific and engineering knowledge
 - ▶ experimental data
- Both can be expressed in narrative form
 - ▶ In fact, humans usually acquire knowledge from narrative forms: lectures, stories.
 - ▶ We have difficulties learning from tables, databases etc.
- There is no reason why the AM-based approach should provide lower predictive power than the model based one.

If we desperately want to match the model based approach:

- assume all stories are relevant
- hide a just-in-time model building algorithm in the interpretation step

Performance

- Can we afford to store the full autobiographical memory of an agent?
 - ▶ We don't want to store "all the data humanity had ever produced (*big data*) but "all the narratives a given human had seen
 - ▶ 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.
- How many stories are relevant at any given time?
 - ▶ An airline pilot is required to have 1500 flight hours.
 - ▶ The experience of a trial lawyer is at most a hundred cases.
 - ★ Of course, these are complemented by books read etc.

The beauty of models

- Still, wouldn't the extracted models be a more compact and elegant representation?
 - ▶ Yes, provided they are compact and elegant (“physics envy”)
- It is not clear that compact models are possible in other fields
 - ▶ Social behavior...

Xapagy cognitive architecture

- Goal: mimicking the ways humans reason about stories
- Stories described in Xapi (“pidgin”) language
- Simple sentences
 - ▶ Subject-verb-object, subject-verb, subject-verb-adjective
 - ▶ Subject-communicative verb-scene + quote (only compound sentence)
- Subjects and objects are represented as **instances**
 - ▶ attributes of instances are represented as overlays of **concepts**
- Sentences mapped to **verb instances** (VIs)
- Newly created VIs are entered into the **focus**.
- During their stay in focus, VIs and instances acquire salience in the **autobiographical memory** (AM).
- VIs are connected by **links** to other VIs present in the focus (succession, elaboration/summarization, context/relation)
- After they expire from the focus, instances or VIs can never return, can never acquire new attributes or links.

The ALIGN step: 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
- Maintenance done by a number of dynamic processes called **diffusion activities** (DAs)
 - ▶ Strengthen VI/Instance shadows based by attribute matches
 - ▶ Scene sharpening
 - ▶ Story consistency
 - ▶ Use probability-proportional-to-size sampling for highly repetitive but low salience events

The EXTEND step: link following

- VIs in the AM are connected using links
 - ▶ succession / precedence
 - ▶ coincidence
 - ▶ context / relation
 - ▶ summarization / elaboration
- The extension of the shadows (matched and aligned stories) into the future is based on a triplet called the Focus-Shadow-Link (FSL) object.

F: "Achilles" / wa_v_sword_penetrate / "Hector".

S: "Mordred" / wa_v_sword_penetrate / "Arthur".

L: "Arthur" / changes / dead.

The INTERPRET step: FSL interpretation

- Source of prediction: the L component
 - ▶ VIs happened in past storylines aligned with the current ones
- Problem: the L components refer to the shadowing storyline!
 - ▶ Predicts the death of Arthur, not of Hector!
 - ▶ So, ok it predicts the death of one combatant, but which one?

- Answer: **reverse shadow**

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

- FSLI (FSL Interpretation) object:
 - ▶ creating all the feasible combinations of interpretations
 - ▶ weighting them according to the ratios in the inverse shadow.

```
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)
```

The INTERPRET step: headless shadows

- 1000s of FSLs → 10,000s of FSLIs
 - ▶ But many FLSs have similar or close interpretations
 - ▶ The number of predictions (with significant weight) are much smaller
 - ▶ Perform similarity clustering over FSLI
- Headless shadow
 - ▶ Clusters of FSLIs with similar interpretation
 - ▶ Looks like a shadow but the head of the shadow is a template **not yet instantiated**
- Combination of supports
 - ▶ Depends on the type of reasoning (continuation, summarization, missing action inference, missing relation inference, question search...)
 - ▶ For continuation:
 - ★ +Succession, +Coincidence, +Elaboration
 - ★ -Shadow, -Predecessor

One-to-one combat domain

- Xapagy 1.0.366 (current 1.0.415 numerical results might slightly differ)
- Domain description: basic + specially designed one-to-one combat domain
 - ▶ concepts and verbs for sword-fight, sword-fencing, boxing
- Synthetic autobiography
 - ▶ series of stories relevant
 - ▶ Hector-Patrocles, Achilles-Pentesilea
 - ▶ King Arthur-Mordred
 - ▶ Cassius-Clay vs. Sonny Liston, Muhammad Ali vs. George Foreman

The duel of Achilles vs Hector

```
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.
```

Shadow maintenance

The shadows of Hector at the end of the story (t=8222)

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)

Shadow maintenance (cont'd)

In Xapagy entities which in colloquial speech are the same might be represented by different instances. The instance of Hector who killed Patrocles is not the same who is fighting with Achilles!

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.

Continuation HLSs

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.

Predicting the outcome

Successively instantiate the strongest HLS.

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.

Predicting an alternative outcome

If we try to find a non-violent end, we 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".
-----
```

Predicting an alternative outcome

8220 "Hector" / wa_v_concedes-defeat / "Achilles".

8221 "Achilles" / wa_v_shakes_hand / "Hector".

8222 "Achilles" / wcr_vr_victorious_over / "Hector".

1 Introduction

2 Implementation

3 Experiments

4 Conclusions

Conclusions



Questions?