Proceedings of the Remembering Who We Are – Human Memory for Artificial Agents Symposium, Mei Yii Lim and Wan Ching Ho (Eds.), at the AISB 2010 convention, 29 March – 1 April 2010, De Montfort University, Leicester, UK

# Forgetting and Generalisation in Memory Modelling for Robot Companions: a Data Mining Approach

**Patricia A. Vargas[1], Alex A. Freitas[2], MeiYii Lim[1], Sibylle Enz[3], Wan Ching Ho[4] and Ruth Aylett[1]**

**Abstract.** The idea of a robot as a long-term companion has not been widely accepted and sometimes not even considered or imagined. Hence, several issues should be addressed in order to facilitate this long-term human-robot interaction. This work focuses on the memory modelling for artificial companions and its forgetting processes by proposing a novel approach based on hierarchical classification data mining techniques to implement memory generalisation mechanisms.

## 1 INTRODUCTION

Modelling a human-like memory has always fascinated AI researchers and has led to a variety of memory models contributing to the understanding of human cognition (Franklin, 1997). In the earlier years, modelling the major characteristics of these memories allowed intelligent programs to remember situations as cases and, most importantly, to extract reasoning rules from these cases stored in a database (see Kolodner (1993) for an overview).

Overall, the organisation of a human memory has a hierarchical structure in the sense that memories of single events are nested within larger cognitive structures. Moreover, most of the events particulars themselves are also organised in a hierarchical fashion (Neisser, 1986; Wright and Nunn, 2000). This intrinsic hierarchical structure seems to permit the storage of an enormous amount of information. In spite of that, if we were to record every bit of incoming information, we would have information overload, difficulty in organizing the information and in focusing on one piece of information at a time. Therefore, forgetting is essential and useful.

Recently there has been an increasing interest in modelling memory for artificial companions, either robots or graphical characters, which could potentially establish long-term relationship with the user (Companions, 2007; LIREC, 2008). The plethora of information that should be "processed" by the robot tends to be enormous each day. Therefore, the generalisation mechanisms involved in such processing is seen as crucial and should be addressed while modelling the memory of long-term robot companions.

Our objective in this work is twofold. First, to propose a means of tackling this generalisation process via a conceptual framework based on data mining techniques for hierarchically structured data, in particular, techniques from the area of Hierarchical Classification (HC) (Sun and Lim, 2001; Freitas and Carvalho, 2007). Second, to *prepare* the robot's memory data (Witten and Frank, 2005) in order to exploit its inherent hierarchical structure. It should be noted that the data preparation is fundamental before applying any HC technique.

To the best of our knowledge, our proposed data mining approach has a novel aspect, as it has not yet been applied to the memory modelling of robots as long-term companions. In general, HC has not yet been applied to any robotics-related problem in the literature. Another contribution of this paper is the proposal of high-level attributes of the memory's data suitable for the generalisation based on HC techniques. These high-level attributes represent a relevant abstraction of the vast amount of low-level data captured about the robot's daily activities and their identification requires expert knowledge about the robot's activities.

Section 2 describes how to prepare the memory data for the memory model together with some relevant psychological background on forgetting and generalisation processes. Section 3 presents a brief account on HC methods. Section 4 describes the proposed conceptual data mining framework for the hierarchical memory and provides an example of the application of our conceptual framework to illustrate our proposal. Section 5 draws some conclusions and suggests future work.[1]

## 2 TOWARDS A HIERARCHICAL DATA STRUCTURE OF THE ROBOT'S MEMORY

The human memory works on the basis of three different processes. First, information from the external and internal sensory system of the organism is encoded. Second, the information is stored either in the short-term memory (STM) or long-term memory (LTM). Finally, the information can be remembered or retrieved (if it has not been forgotten). Note that it is as yet unclear and controversial among scientists, how exactly human memory works (Wilson, 1994; De Zeeuw, 2005; Levenson, 2006).

According to one of the existing models of the human memory (Tulving, 1972), LTM can be divided into declarative and procedural memory. The declarative memory is formed of a semantic and an episodic memory. Autobiographic memory (AM), which is the focus of our research, is a specific kind of episodic memory that contains relevant and meaningful personal experiences for a human being (Conway, 1990; Nelson, 1993).

In our recent work (Ho et al., 2009; Lim et al., 2009) we have proposed an autobiographic memory model for a long-term

[1] Dept. of Computing Science, Heriot-Watt University, Edinburgh, Scotland, UK. Email: {p.a.vargas,myl,ruth}@hw.ac.uk
[2] Computer Science Department, University of Kent, Canterbury, UK. Email: a.a.freitas@kent.ac.uk
[3] Otto-Friedrich Universitaet Bamberg, Germany. Email: sibylle.enz@uni-bamberg.de
[4] Dept. of Computing Science, University of Hertfordshire, England, UK. Email: w.c.ho@herts.ac.uk

companion whose information stored in the long-term episodic memory component of the memory architecture was modelled as a collection of events. Nonetheless, there was no hierarchical structure reflected in the data stored.

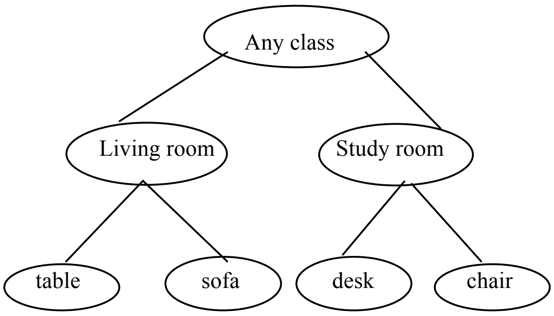


**Fig. 1**. Class hierarchy example for the memory's data.

Therefore, one of our aims in the present work is to analyse the information to be stored in the robot's memory and compose a novel hierarchical structure. Our enhanced memory model will enable the companion to remember events that are relevant or significant to itself or to the user. For other events with a lower long-term value, the memory model will support forgetting through the processes of generalisation and memory restructuring. The long-term memory module in particular, contains episodic events that are chronologically sequenced and derived from an agent's interaction history both with the environment and the user. Meanwhile LTM also produces concepts as knowledge about the world in order to help in formulating and processing new goals.

We propose a novel structure taking into consideration its inherently hierarchical nature (Fig. 1). Basically, the autobiographic LTM will be composed of events/episodes (or data instances) and the information related to each event will be stored and organised as attributes. It is beyond the scope of this paper to describe how the data will be collected and stored following the proposed structure. Our focus here is on the proposed generalisation mechanisms to be implemented.

Table 1 enumerates the key attributes that should be part of the robot's memory every time it records an event. Each event should be decomposed into the *subject* "who" or *person* of the *action*, the *object* "whom" or *person* directly involved in the *action* and the *action* "what" itself. The *time* "when" (has a hierarchical structure) and place or *location* "where" the *action* took place (has also a hierarchical structure) should also be part of the event particulars. Finally, the robot should store its "*internal state"* related to the event, e.g. its emotional status, and the subject's internal state, e.g. her/his emotional status, named here "external state".

Table 1. Attributes of each event to be stored.

| ID | NAME | DESCRIPTION | TYPE |
|---|---|---|---|
| 1 | Who | Subject/Person | Alphabetic |
| 2 | Whom | Object/Person | Alphabetic |
| 3 | What | Action | Alphabetic |
| 4 | Where | Location | Alphabetic |
| 5 | When | Time | Numeric |
| 6 | Internal State | Robot's emotional status | Alphabetic |
| 7 | External State | Whom's emotional status | Alphabetic |

# 3 HIERARCHICAL CLASSIFICATION (HC)

Classification is one of the most important problems in data mining (Witten and Frank, 2005). Basically, it can be defined as (Freitas and Carvalho, 2007): given a set of training instances (examples) composed of pairs $\{x_i, y_i\}$, find a function *f(x)* that maps each attribute vector $x_i$ to its associated class $y_i$, $i = 1, 2, \ldots, n$, where $n$ is the total number of training instances.

After training, the predictive accuracy of the classification function induced is evaluated by using it to classify a set of unlabeled (unknown-class) instances, unseen during training. This evaluation measures the generalisation ability (predictive accuracy) of the classification function induced.

In Hierarchical Classification (HC) problems (Sun and Lim, 2001; Freitas and Carvalho, 2007; Vens et al., 2008), the classes are disposed in a hierarchical structure, such as a tree or a direct acyclic graph. In these structures, the nodes represent classes, and the edges typically represent a "is-a" relationship between classes. If an instance is assigned to a given class *c* that instance should also be assigned to all of the ancestral classes of *c*.

We emphasise that in general what defines a HC task is the fact that the class attribute has hierarchically-structured values, regardless of whether the other attributes (the predictor attributes) have hierarchically-structured values or not. Our problem is even more complex than a conventional hierarchical problem for in our case not only the class attribute has a hierarchical structure but also some predictor attributes have hierarchically-structured values.

There are two major types of HC approaches, namely the "top-down" approach and the "big-bang" approach (Witten and Frank, 2005). In essence, in the top-down approach a set of classification models is built from the training set and their predictions are combined in a top-down fashion. The classifier predicts, for the current instance, one of classifier's child classes at level 1. The instance is further passed down to the classifier whose child class was just predicted, and so on until the instance cannot be further classified (e.g. when a leaf class node is reached). An example of this top-down HC process will be discussed, in the context of the proposed memory's data, in the next section.

In the big-bang approach, a single classification model is induced which is in principle capable of predicting any class, at any level of the hierarchy, for a test instance. HC algorithms following the big-bang approach in general are more difficult to design and implement than algorithms following the top-down approach, since in the latter each of the classifiers can be induced in a modular fashion, solving a simpler problem, therefore implementing a kind of "divide and conquer" approach.

In this work we consider initially the top-down approach for HC, given the complexity of the target problem and the lack of previous work in the area, it seems sensible to start with a relative simpler approach before trying more complicated ones.

# 4 THE "DREAMING PROCESS"

Memory forgetting is useful to improve efficiency, scalability and adaptability of cognitive systems operating in dynamic environments, such as a robot's interaction environments. Forgetting could be viewed as a way of controlling the memory of the robot companion for it could be used to *regulate* the type

Table 2. Attributes of each event to be stored (a small subset of the entire dataset).

| Event | Who | Whom | What | Where | When | Internal State | External State |
|---|---|---|---|---|---|---|---|
| 1 | Amy | Apple | Eat | LivingRoom.Table | 2009.06.11.Morning | Happy | Happy |
| 2 | Claire | Game | Play | Living Room.Sofa | 2009.06.11.Evening | Happy | Excited |
| 3 | Amy | Game | Play | Living Room.Sofa | 2009.06.12.Evening | Happy | Excited |
| 4 | John | Book | Read | StudyRoom.Desk | 2009.06.13.Afternoon | Default | Sleepy |
| 5 | John | Andy | Chat | StudyRoom.Chair | 2009.06.13.Evening | Happy | Happy |
| 6 | John | TV | Watch | StudyRoom.Desk | 2009.06.14.Evening | Concerned | Bored |

and amount of data stored on the robots memory (Gold,1992), giving rise to a more reliable companion, in terms of data security and privacy (Vargas et al., 2009; Lim et al., 2009).

Neuroscientists and psychologists have developed a number of theories of forgetting including generalisation processes. Generalisation describes processes of abstraction in the human memory. Concrete episodes are continuously restructured loosing their concrete features and being reduced to their core meaning, forming abstract categories that include a variety of episodes with different concrete features. This process serves the limited cognitive capacity of the human brain and allows for quick and easy categorization and representation of a complex, ever-changing environment. Generalisation serves the successful use of memory in daily life as it allows for abstraction from concrete and singular experiences to develop a model of the extremely complex and ever changing environment. Hence, here we propose a mechanism named "dreaming process" to implement a generalisation process in the robot companion's long-term memory.

In our proposed long-term memory model the attributes (including the class attribute) of the events or episodes (data instances) are going to be recorded and stored using a hierarchical data structure, hence, particularly suitable for HC techniques (Section 3). Initially, we plan to use the dreaming process to "clean" the autobiographic long-term memory by means of generalisation. Nevertheless, it important to emphasize that this process will also be responsible for generating a knowledge base representation that will enable the companion to learn and make "predictions" if required.

Basically, the information that receives frequent attention will go through reconstruction processes before it is consolidated as LTM. By being able to notice and recall differences in experiences, the robot companion will be able to learn about its environment more effectively. General structures will help the companion in deciding what to pay attention to, and "reminding" forces it to make use of prior knowledge to form expectations. Nonetheless, care needs to be taken when generalising information to ensure that particular differences that may be valuable are not lost. This anticipated "dreaming process" is crucial for the maintenance of a meaningful autobiographic long-term memory in the sense that only relevant events will be kept stored in the memory database and upcoming information will be stored according to its relevance and pertinence to previously stored data.

### 4.1 Hierarchical Classification in Memory Modelling

Our proposed conceptual framework for HC is based on the top-down HC approach (Section 3). The proposed HC method will run during the "dreaming process" of the robot, which could take place every night, or at any predetermined time slot. In order to define our target HC problem, one must first define the instances, the predictor attributes and the class attribute.

Table 2 depicts an example list of events or data instances (a small subset of the entire dataset) that are being stored by a robot. In this paper we are rather interested in the HC problem that arises when the attribute chosen as the class attribute is a hierarchically-structured one. Suppose that the attribute "where" of Table 2 is chosen as the class attribute. Assuming the robot is moving just inside a house, the first data instance shows that the event occurred in the general location *living room*, and more specifically in the location *table*. In this example scenario, the class hierarchy is show in Figure 1. In order to implement the proposed top-down HC approach, we would need to train a classifier for the root node that will discriminate between classes *living room* and *study room*, a classifier for the *living room* node will discriminate between subclasses *table* and *sofa*, and a classifier for the *study room* node will discriminate between subclasses *desk* and *chair*. In the testing phase, each test instance is first assigned to the root classifier.

We will show how the generic method can be implemented by using a simple Naïve Bayes classifier (Witten and Frank, 2005) as the base classifier. Naïve Bayes is a well-known classification algorithm. We are using it here for its conceptual simplicity.

We show here a subset of the probability calculations associated with the training of Naïve Bayes at the root node, a conceptually similar set of probability calculations would be performed for the Naïve Bayes trained at the class nodes *living room* and *study room*. Let "Where" be the class attribute to be predicted and let $X_i$, $i$ = 1,.,,6, denote the predictor attributes Who, Whom, What, When, Internal State, External State, respectively. Naïve Bayes has to compute, for each class, the value of its posterior probability P(class|attributes) = P(attributes|class) × P(class) = $\prod_i$ P($X_i$|class) × P(class) (1)
where P(class) is the class's prior probability, and then it assigns to an instance to be classified the class with the largest posterior probability value.

Using the very small dataset in Table 2, the prior class probabilities for the classifier at the root node would be: P(*living room*) = 3/6 = 1/2, P(*study room*) = 3/6 = 1/2. The conditional probabilities of the form P($X_1$|class), for example, would be computed as P(*Who = Amy|living room*) = 2/3, P(*Who = Claire|living room*) = 1/3, P(*Who = John|living room*) = 0/3 = 0, P(*Who = Amy|study room*) = 0/3 = 0, P(*Who = Claire| study room*) = 0/3 = 0, P(*Who = John| study room*) = 3/3 = 1. The conditional probabilities of the form P($X_i$|class) for $i$ = 2,…6 can be computed in an analogous way.

Suppose now that a given test instance is submitted to the Naïve Bayes classifier trained for the root node. After considering the values computed by Equation 1, for each class, as a function of the attribute values observed in the test instance, the classifier would assign the class *living room* or *study room* to the instance. Note that, if the test instance has the value *John* for the attribute *Who*, this fact by itself would suggest (ignoring other attributes for now) that the most likely location for an event involving *John* (at the first level of the class hierarchy) would be the *study room*, since P(*Who* = *John*| *study room*) = 1 and P(*Who* = *John*|*living room*) = 0 and the two classes have the same prior probability (1/2). Of course, the prediction of the location of an event involving John will also consider all the other predictor attributes too (calculations not shown due to lack of space). So, if the user wants to predict the location of an event involving John, the user will have to provide all the contextual information for the event, such as whom John would be interacting with, what John would be doing, etc. A similar prediction process would be performed for the most specific level of the location class hierarchy, as per the above-described top-down approach.

## 5 DISCUSSION AND FUTURE WORK

In this work we propose a novel approach towards the goal of devising an enhanced memory model for a long-term robot companion including the use of HC data mining techniques to implement generalisation mechanisms. We believe that generalisation is useful to improve efficiency, scalability and adaptability of cognitive systems operating in dynamic environments, such as a robot's interaction environments. We plan to address all forgetting mechanisms in our robot's memory model (Vargas et al., 2009; Ho et al., 2009, Lim et al., 2009) and we see this present work as a step forward towards this endeavour.

We first propose to structure the memory data hierarchically for it is envisaged that this hierarchical data structure will facilitate the implementation of the aforementioned forgetting and generalisation mechanisms. We have also provided a simple example of how the proposed conceptual framework could be applied to the memory generalisation process using a HC method.

HC concepts and techniques are particular suitable to our memory model for the attributes (including class attributes) of the memory's events or episodes (represented as data instances) will be prepared to be recorded and stored using a hierarchical data structure. To the best of our knowledge, computational models of AM have neither so far adequately accounted for its hierarchical organisation and nested structure nor had fully exploited these characteristics (Ho et al., 2008).

We are currently working in the implementation of the proposed conceptual framework based on HC, initially using a Naïve Bayes algorithm as the base classifier, given its simplicity, computational efficiency and relatively good predictive performance, and we will also investigate more sophisticated kinds of base classifiers in the future.

**Acknowledgments.** This work was partially supported by European Community (EC) and is currently funded by the EU FP7 ICT- 215554 project LIREC (LIving with Robots and Interactive Companions). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

## REFERENCES

Companions (2007). http://www.companions-project.org/

Conway, M. A. (1990), Autobiographical Memory: An Introduction. Buckingham: Open Univ. Press.

De Zeeuw, C. I. (2005). "Time and tide in cerebellar memory formation," Current Opinion on Neurobiology, vol. 15, pp. 667–674.

Franklin, S. (1997). "Autonomous agents as embodied ai," Cybernetics and Systems Special Issue on Epist. Aspects of Embodied AI, vol. 28, no. 6, pp. 499–520.

Freitas, A. A. and de Carvalho, A. (2007). A Tutorial on Hierarchical Classification with Applications in Bioinformatics. In: D. Taniar (Ed.) Research and Trends in Data Mining Technologies and Applications, pp. 175-208. Idea Group.

Freud, S. (1937). The Ego and the Mechanisms of Defence. London: Hogarth Press and Institute of Psycho-Analysis, 1937.

Gold, P. E. (1992), A proposed neurobiological basis for regulating memory storage for significant events. Affect and Accuracy in Recall: Studies of 'Flashbulb' Memories, 141-161.

Ho, W. C., Dautenhahn, K. and Nehaniv, C. L. (2008) "Computational memory architectures for autobiographic agents interacting in a complex virtual environment: a working model", Connection Science, 20:1, 21 - 65.

Ho, W. C., Lim, M., Vargas, P. A., Enz, S., Dautenhahn, K. and Aylett, R. (2009), "An Initial Memory Model for Virtual and Robot Companions Supporting Migration and Long-term Interaction", RO-MAN'2009, Japan.

Kolodner, J. L. (1993) Case-Based Reasoning. San Mateo, CA: Morgan Kaufmann.

Levenson, J. M. (2006). Epigenetic mechanisms: a common theme in vertebrate and invertebrate memory formation. Cellular and Molecular Life Sciences, 63:1009-1016.

Lim, M., Ho, W. C., Vargas, P. A., Enz, S. and Aylett, R., (2009), "A Socially-Aware Memory for Companion Agents", 9th International Conference on Intelligent Virtual Agents (IVA'2009), Amsterdan.

LIREC (2008). The LIREC Project, http://www.lirec.org/

Neisser, U. (1986) Nested structure in autobiographical memory, in D. C. Rubin (ed.), Autobiographical Memory, Cambridge University Press, UK, pp. 71–88.

Nelson, K. (1993). "The psychological and social origins of autobiographical memory," Psych. Science, vol. 4, pp. 7–14, 1993.

Sun, A. and Lim, E. P. (2001). Hierarchical text classification and evaluation. In Proceedings of the 1st IEEE International Conference on Data Mining. IEEE Computer Society Press (pp. 521-528).

Tulving, E. (1972). "Episodic and semantic memory," in Organization of Memory, E. Tulving and W. Donaldson, Eds. New York: Academic Press, pp. 89–101.

Vargas, P. A., Ho, W. C., Lim, M., Enz, S. and Aylett, R., (2009), "To Forget or Not to Forget: Towards a Roboethical Memory Control", Proceedings of the AISB-2009, Endinburgh, UK, pp. 18-23.

Vens, C., Struyf, J., Schietgat, L., Dzeroski, S. and Blockeel, H. (2008). Decision trees for hierarchical multi-label classification. Machine Learning 73(2), pp.185-214.

Wilson, M. (1994). Reactivation of hippocampal ensemble memories during sleep. Science, 265.

Witten, I. H. and Frank, E. (2005). Data Mining, practical machine learning and techniques. 2nd ed., Elsevier, Morgan Kaufmann.

Wright, D. B., & Nunn, J. A. (2000). "Similarities within event clusters in autobiographical memory". Applied Cognitive Psychology, 14, 479-489.