Emotions in NARS

Pei Wang
Department of Computer and Information Sciences
Temple University
http://www.cis.temple.edu/∼pwang/

October 1, 2015

Abstract

This memorandum explains the conceptual design of the components of NARS that are directly related to emotion.

1 Introduction

In biological systems, emotion is closely associated with drives like survival and reproduction, according to which decisions are made. In a computer system, there is no biological drive, and the primary driving force is the task that is assigned to the system by the designer or the user.

Traditional AI systems do not need emotion, since there is little need to choose among tasks, which are assumed to be consistent, and within the system's capability. Since NARS is designed under AIKR, the traditional assumption is no longer true, and the system does need to choose among conflicting tasks, as well as to make quick and flexible responses to the environment.

In NARS, the main function of emotion is the appraisal of the external and internal environment with respect to the system's tasks, so as to act accordingly, especially in decision making and resource allocation. Emotion is not an independent process or module in NARS, but is embedded in various places, and tightly entangled with the reasoning/learning processes in the system.

NARS makes no attempt to accurately simulate human emotions. Even so, some resemblance with human will be there, given the functional similarity of the systems.

2 Desirability of Events

Since the overall objective of NARS is to successfully carry out its tasks, and especially to achieves its goals, appraisals are initially start from the goals.

As defined in [Wang, 2013], a goal in NARS is a statement to be realized by the system. Each goal has a desire-value associated, which is defined as the truth-value of the implication statement stating that the realizing of the goal will lead to a (unspecified) desired state. The desire-values of given goals are determined by the user of the system. These goals can be implanted in the system's memory or entered at the user interface.[1]

The derivation and revision of goals are carried out by NAL inference rules, which also calculate desire-values for derived goals. Derived goals are basically handled in the same way as given goals.

Each concept named by a statement will either contain or link to the goals that with this statement as content, and therefore has a desire-value that summarize those of the goals. At the level of the concept as a whole, the summarized "general desire-value" and the "current desire-value" decided by the most active goal at the moment are not necessarily the same.

It is the current desire-value of a concept that decides whether the system will actually attempt to achieve the corresponding goal. Each time a new goal appears (either input or derived), the current desire-value will be updated using the revision rule. After that, the decision rule will compare the new desire-value to a threshold to decide whether to immediately pursue it. If the statement corresponds to an operation, to pursue it means to execute the operation, either by the system itself or by a plug-in tool. If it is not an operation, the system will first make a "reality check" to see whether it is already true, then use the difference between desire and reality as the priority value for the goal to be used in inference to generate derived goals.

It is the last step in the above description where emotional factors are introduced into the system. A "satisfaction" value is defined to be the (absolute) difference between the expectation of the (current) desire-value of the statement and the expectation of the (current) truth-value of the belief on the same statement, so it will be in [0, 1], where 1 means the system gets what it desires, and 0 the opposite. If the statement is an operation, the satisfaction value is obtained from the feedback of the execution, which indicates whether the execution is successful.

A satisfaction value can also be defined if the task is not a goal, but a question or a judgment. Usually the satisfaction is the quality of the answer or the derived judgment, though if the inference is a revision of two judgments, we may want to use satisfaction to indicate the level of consistence between the two (1 minus the difference of their frequency values), so 1 means consistency, and 0 means contradiction.

The satisfaction value is used mainly in two places. Within the concept, it is used to adjust the budget of the task (to decrease its priority according to satisfaction) and belief used to generate the task (to increase its durability according to satisfaction). It is also used to calculate the system-level satisfaction, to be discussed in the following.

[1]The design of NARS makes no assumption about the content or desire-value of the given goals. How to choose proper given goals for each application is a problem to be solved in the future by the people responsible for the application.

3 Feelings of the System

As described above, each working cycle reports a satisfaction value in [0, 1], to indicate the extent the system is "satisfied" in that step.

To get a measurement as the system's appraisal of the current and recent situations, the satisfaction values from each cycle are accumulated into an overall satisfaction. After each cycle, the satisfaction value S is updated to $rs+(1-r)S$, where s is the satisfaction value of the cycle, and r is a system parameter deciding the relative weight of the two factors. Also, s or r may need to be modified according to the importance of the corresponding task.

The current satisfaction value can be brought into the system's experience by a mental operator "$feel$". The operation $feel(SATISFIED)$ can generate an event "$< (\{SELF\} \times [SATISFIED]) \rightarrow feel > . :|:$" with the satisfaction value as frequency (plus a default confidence). This operation can be explicitly invoked as a goal, or automatically triggered when the satisfaction is beyond the neutral zone (around 0.5, defined by a system parameter). In this way, the appraisal enters the system consciousness, and is processed together with the system's other experienced events. Beside the constant term $SELF$ that has been implemented to refer to the system itself, constant term $SATISFIED$ is introduced for a specific feeling. An alternative implementation is to define different mental operators for different feelings. In that way the operation will be $feel\text{-}satisfied()$ and the event be "$< \{SELF\} \rightarrow feel\text{-}satisfied > . :|:$"

Beside **satisfaction**, the system may have a few other system-level indicators that can be felt in this way:

Alertness that summarizes the average difference between recently processed input and the corresponding anticipation.

Busyness that summarizes the average priority values of the recently processed tasks.

Well-being that summarizes the situation of energy supply, I/O channel connection, device functioning, etc.

In a narrow sense these indicators are not emotions, since they are not handled exactly as described in the previous section, though in a similar way. All the corresponding feeling operators can be invoked or triggered as the previous one, and the events generated appear in the system's internal experience.

The *internal experience* of the system is a special I/O channel that consists of the following two types of items: the feelings described above, and the internal operations executed by the system. Conceptually speaking, the former is input, while the latter is output. However, technically they are both operations executed within the system. Here "operations" include the mental operations defined in NAL-9, as well as the inference operations carried out, which will be recorded as implication judgments between the task as premise and the task as conclusion, while the belief used in the inference will be omitted as background knowledge. The truth-value of the implication will be calculated as induction.

The internal experience may be treated similarly as the other I/O channels, with a buffer and other features. When this channel is merged into the overall experience channel, that channel will include both external and internal experience, and therefore represents the system's consciousness.

4 Emotion in Concepts

The concepts in NARS can be "emotional" in two senses.

As described previously, every concept corresponding to statement gets a general desire-value from the goals with the statement as content. Now this value is added to all concepts, and at the end of each working cycle, the desire-value of the fired concept is adjusted by the current satisfaction value, again using a weighted average. So, in the long run, the desire-value of a concept indicates the extent to which its firing is associated with high satisfaction of the system, which provides an appraisal for the relation between the concept and the goals.

To bring this appraisal into the internal experience of the system, a feeling operator can be invoked with a term as argument to generate an event indicating how much the system "likes" (or "dislikes") the term. This operator can also be triggered by an extreme (high or low) desire-value in the concept.

Beside this "emotional indicator" in every concept, there are also special concepts whose meaning is especially emotional. The basic concepts in this group include all the feeling constants like $SATISFIED$. These concepts provide the building blocks for the system's feelings and emotions.

Starting from the basic feelings, more complicated feelings can be built by combining them and the other concepts. For example, an event with the same desire-value may become different feelings when combined with other features, such as "it has happened" vs. "it will happen", "it is caused by the system itself" vs. "it is caused by someone else", "it is manageable" vs. "it is inevitable", etc. The new feelings are formed using the same composing rules as other compound terms, and their generation is experience-driven. These compound feelings may or may not correspond to human feelings.

5 Effects of Emotion

As described above, in NARS emotional information appears at two separate levels:

- in the "subconscious level", as desire-values of concepts or attributes of memory,
- in the "conscious level", as terms and statements, expressed in Narsese.

Emotions in both senses will contribute to the system's behaviors.

An important usage of these "emotional concepts" is to categorize the current situations, as well as to develop strategies to deal with such situations. For instance, there may be many very different situations that can be categorized as "dangerous", so as to be handled with some common responses. In this aspect, their treatment is similar to that of the other concepts.

The "emotion-specific" treatments will mainly happen in the "subconscious level", where the emotional information is used in various processes, such as

- The desire-values of concept will be taken into account by the budget functions, where concepts with strong feeling (extreme desire-values) will get more resources than those with weak feeling (neutral desire-values).
- In the decision-making rule, the threshold for a decision will be lower in high emotional situations, so has to allow quick responses.
- The overall satisfaction will be used as feedback to adjust the desire-values of data items (concepts, tasks, beliefs), so that the ones associated with positive feeling will be rewarded, and the ones associated with negative feeling punished. In this way, the system will show a "pleasure seeking" tendency, and its extent can be adjusted by a system parameter.
- When the system is busy, tasks with low resource budget will be simply ignored to achieve high efficiency. This attribute value can be used as a parameter in the priority–probability mapping to control the "level of focus" of the system's attention.
- When the system is alert, it will spend more time to process new tasks in the input buffer, which means less time to the existing tasks in memory.
- When the system "does not feel well", it will spend more time in the related self-maintenance tasks, which means less time for other tasks.

The above list is by no means complete. When the control mechanism of NARS is revised, emotional factors will be considered in all steps.

Emotion also plays an important role in communication with other systems. However, unlike many "affective computing" projects, in NARS this function of emotion is secondary, and depends on its primary function in self-control. How to use emotion to improve cooperation with other systems is a topic to be studied in detail in the future. At the current stage, some displays can be added in the GUI to show the related information, such as the satisfaction level of the system or the desire-value of concepts.

References

[Wang, 2013] Wang, P. (2013). *Non-Axiomatic Logic: A Model of Intelligent Reasoning*. World Scientific, Singapore.