Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 38 (2011) 4447-4450

ISSN: 2229-712X

Soft computing in intelligent multi – modal systems

J. Deepa Sajani

Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai - 625009.

ARTICLE INFO

Article history:
Received: 22 August 2011;
Received in revised form:
26 August 2011;
Accepted: 31 August 2011;

Keywords
Instrumentation,
Multi-modal systems,
Non intrusive neural network based
Gaze tracking system,
Reasoner using fuzzy and probabilistic technique.

ABSTRACT

As man has started using computers to track the work of machines, the flexibility with which the humans communicate and request information from the computers must increase. There arises a need for the computers to understand and process the ambiguous queries put forth by authorized persons. This has been accomplished in this paper by introducing soft computing in multi – modal systems. This architecture first uses vascular pattern recognition to identify whether the user is an authorized person or not by using the image of his blood vessels obtained by using near – infrared light. Then the architecture consists of six main components - a reasoner, a speech system, a vision system, a non intrusive neural network based gaze tracking system, an integration platform and an application interface, using which we can process speech, text, face images, gaze information and simulated gestures. Fuzzy and probabilistic techniques have been used in the reasoner to establish temporal relationships and learn interaction sequences. This architecture can be used in Process control and Instrumentation where the authorized users can query the values of some physical parameters (pressure, temperature etc). The ambiguities of these queries are resolved using gaze tracking.

Example:

Query (recognized by speech): show me this of that machine. (That machine is pointed using mouse) Ambiguity: this (physical parameter) - identified by where the user's gaze is fixed on the screen. Final query: show parameterA of machineB.

Introduction

Soft computing technique is a natural way of handling the flexibility with which the humans communicate, describe the events, and perform actions or request information from the computer. Till today human – computer interaction has mainly focused on natural language, text, speech and vision in isolation. There arises a need of integrating all these modalities using intelligent reasoners so that the ambiguities in single modes of communication can be resolved if extra information is available. An Intelligent multi – modal system combines, reasons with and learns from information, originated from different modes of communication between human and the computer.

In this paper an intelligent multi – modal system has been developed which can be used in process control and instrumentation to easily query the different physical parameters of the machines.

User Authentication

Initially, the computer identifies only the authorized person and allows only them to access the information of the machines. Several technologies like finger imaging, facial recognition, voice recognition, signature recognition are prone to a large number of errors and hackings. So, in our system we employ the VASCULAR PATTERN RECOGNITION technology. This is also known as the palm vein technology.

Technology:

- The user first holds his palm a few centimeters above the palm vein scanner.
- The sensor passes a near infrared ray on the palm.
- Deoxygenated hemoglobin in the blood, flowing through the veins absorbs the near infra - red rays, illuminating the hemoglobin, causing it to be visible by the scanner.
- Arteries and capillaries, whose blood contains oxygenated hemoglobin, which does not absorb near-infrared light, are invisible to the sensor.
- The still image captured by the camera, which photographs in the near-infrared range, appears as a black network, reflecting the palm's vein pattern against the lighter background of the palm.
- An individual's palm vein image is converted by algorithms into data points, which is then compressed, encrypted, and stored by the software
- Each time a person logs in attempting to gain access by a palm scan to the computer; the newly captured image is likewise processed and compared to the registered one.
- Numbers and positions of veins and their crossing points are all compared and, depending on verification, the person is either granted or denied access.

Fig.1. Palm on sensor

Tele:
E-mail addresses: sajanitce@gmail.com

Fig.2. Process of Registration

Advantages:

- Vein recognition technology is secure because the authentication data exists inside the body and is therefore very difficult to forge.
- It is highly accurate.
- No two people in the world share a palm vein pattern - even those of identical twins differ.
- If you registered your profile as a child , it'll still be recognized as you grow, as an individual's patterns of veins are established in utero (before birth).
- The completely contactless feature of this device makes it suitable for use where high levels of hygiene are required.

Input and Output

Intelligent multi-modal systems use a number of input or output modalities to communicate with the user, exhibiting some form of intelligent behavior in a particular domain. The functional requirements of such systems include the ability to receive and process user input in various forms such as:

- typed text from keyboard,
- mouse movement or clicking,
- speech from a microphone,
- Focus of attention of human eye captured by a camera.

The system must be also able to generate output for the user using speech, graphics, and text.

Multi – modal system function

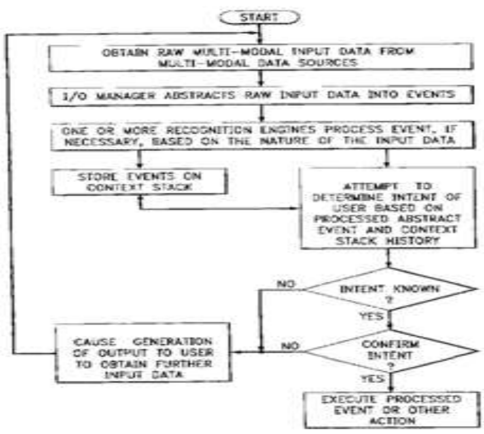


Fig.3. Flow Chart of functionality of Multi-Modal System Architecture

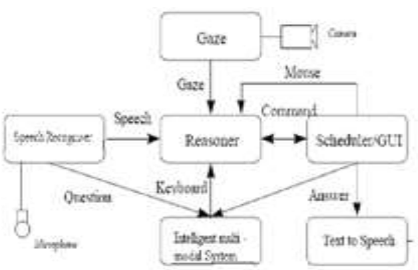


Fig.4, Architecture of Multi – Modal System

The Reasoner

Functions:

- It must be able to handle ambiguities such as give me this of that. The reasoner handles ambiguities by using a knowledge base that is being continually updated by the information arriving from various modalities.
- It must have the capabilities to deal with often-conflicting information arriving from various modalities.

Structure:

The reasoner has five modules. They are:

- Fuzzy temporal reasoning,
- Query pre-processing,
- Constraint checking,
- Resolving ambiguities (WIZARD)
- Post-processing.

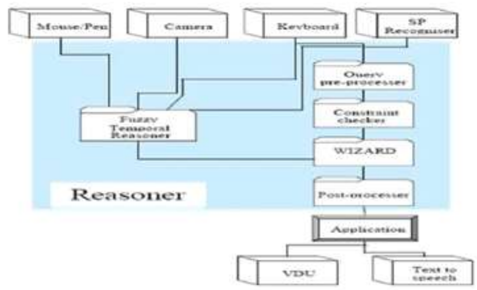


Fig.5, Structure of a Reasoner

Fuzzy temporal reasoning module:

Method:

Time-stamped events from various modalities are received and the fuzzy temporal relationship between them is determined. Input can be defined by their preposition and time quantifier. Preposition describes the content of the input, and the time quantifier is the time interval of duration of the alarm. We will refer to the time of origination of an alarm as $\boldsymbol{T_b}$ and the time of termination as $\boldsymbol{T_e}$.

Temporal reasoning i.e. reasoning about time, plays a critical role. The system should be able to reason about the relative and absolute time of occurrence of events, duration of events, and sequence of events. The temporal relation between any two events can be determined by comparing their origination and termination time (T_b and T_e).The comparison is to check the relation between these times (Less, Equal or Greater).Fuzzy logic is recommended for finding the temporal relation between events for correlation purpose.

Let e_1 and e_2 be two events defined on an interval time[T_{1b},T_{1e}] and [T_{2b},T_{2e}] respectively.Table1 shows the different possible temporal status of $_{e1}$ relative to $_{e2}$ where ~>,~<and ~= are the fuzzy symbols for greater, less and equal respectively

In the proposed model four input linguistic variables are defined to represent the temporal differences $(T_{1b} - T_{2b})$, $(T_{1e} - T_{2e})$, $(T_{1b} - T_{2e})$, and $(T_{1e} - T_{2b})$, and one output linguistic variable - *Status* - to represent the temporal relation between the two events.

The relationship with the highest status will be chosen as the most likely relationship between the two events. This relationship can be used later by the reasoner to resolve conflicts between, and checking dependency of the modalities.

Advantages:

The advantage of fuzzy temporal reasoning over the traditional crisp reasoning is given in table 2.

Table I Temporal relations

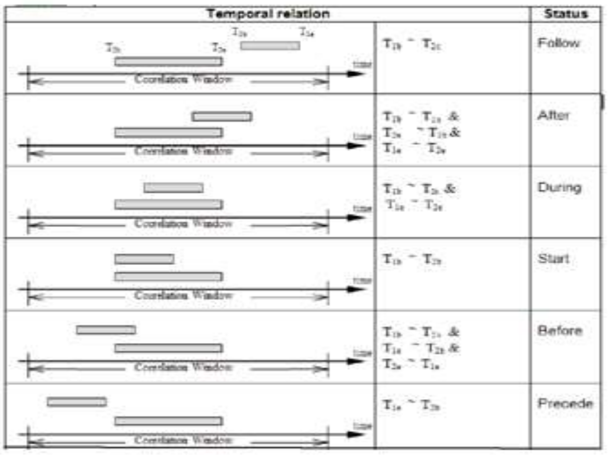
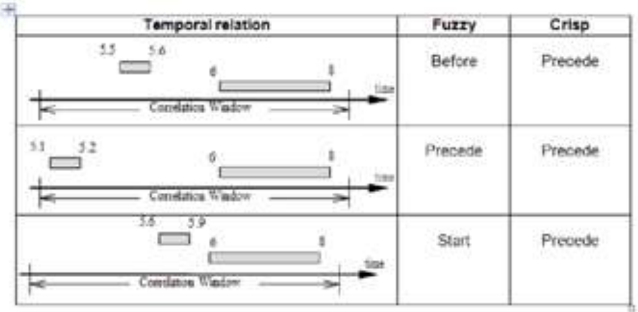


Table III Advantage of fuzzy reasoning



The first event starts before the second one at different times. In all cases the crisp model considers the temporal relation between both events as "Precede" (the first event precedes the second one.) The fuzzy model makes different decisions according to how long the period between the first and second event is.

Query pre-processing module:

It converts a sentence in natural language form to a query which conforms to the system's pre-defined grammar. Redundant words are removed, keywords are placed in the right order and multiple word attributes are converted into single strings.

Constraint checking module:

It examines the content of the queries. If individual parts of the query do not satisfy pre-defined constraints then they are replaced by ambiguous terms (this, that) to be resolved later, otherwise the query is passed on to the next module.

Wizard:

This is at the heart of the reasoner and is the module that resolves ambiguities. The ambiguities in this application take the form of reserved words such as this or that, and they refer to objects that the user is or has been talking about, pointing at or looking at. The ambiguities are resolved in a hierarchical manner.

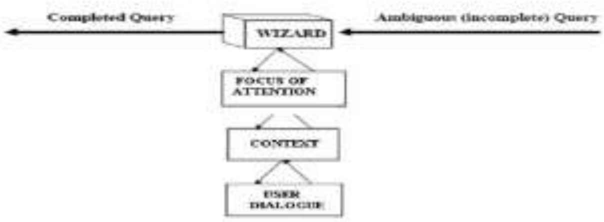


Fig.6, Resolving ambiguities in a Reasoner

The context of the interactions between the user and the system, if it exists, is maintained by the reasoner in the knowledge base. When a new query is initiated by the user, it is checked against the context. If there are ambiguities in the query and the context contains relevant information then the context will be used to create a complete query that will be sent to the application interface for processing.

Another mechanism used to resolve ambiguities is the focus of attention, which is obtained from the user when pointing with the mouse or gazing at an object on the screen. At the same time there could be dialogue with the user through the speech recognizer or the keyboard.

All the modules work in parallel. This can cause conflict between information arriving from various modalities. The conflict resolution strategy used in the reasoner is hierarchical with the focus of attention having the highest priority and the dialogue system the lowest. This means that the dialogue system will act as a safety net for the other modalities if all fails, or if inconsistent information is received from the modalities. In cases where text input is required however, the dialogue system is the only modality that will be called upon. In all other cases the dialogue system will be redundant unless all others fail in which case a simple dialogue in the form of direct questions or answers will be initiated by the system. The wizard sends the completed queries to the post-processing module.

Post-processing module:

It simply converts the completed queries in a form suitable for the application. This involves simple operations such as formatting the query or extracting key words from it.

Table IIII
Examples of interaction with the reasoner

Query	User Action	Reasoning Process
Show me machine A location on the Map	None	Complete query, Process command. Sent to application.
Show me this of that machine	Focus on machine B Mouse points to pressure	Two ambiguities. Resolved.
Read me this	No focus No context	Every thing is ambiguous. User is asked to repeat the missing parts. Context is updated.

Gaze Tracking

The initial objective for the gaze system is the capability of tracking the eye movements within slightly restricted environments. More specifically, the scenario is a user working in front of a computer screen viewing objects on display while a camera is pointed at the user's face. The objective is to find out where (which object) on the screen the user is looking at, or to what context he is paying attention. This information in combination with other inputs provided by speech and other modalities would be most useful to resolve some real application tasks.

System Description:

The approach is to detect and track different facial features, using varying image analysis techniques. These features will serve as inputs to a neural net, which will be trained with a set of predetermined gaze tracking series. The output is coordinates on the screen.

Neural networks have been chosen as the core technique to implement the gaze tracking system.

• A three-layer feed-forward neural network is used. The net has 600 input units, one divided hidden layer of 8 hyperbolic units each and corresponding divided output layer with 40 and 30 units, respectively, to indicate the positions along x- and y- direction of the screen grid.

• Grey-scale images of the eye are automatically segmented from the head images inside a search window, the images of size 40*15 are then normalised, and a value between -1 and 1 is obtained for each pixel. Each normalised image comprises the input of a training pattern.

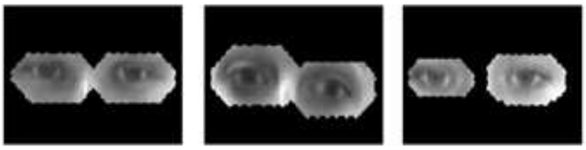


Fig.7, Grey scale images of eye

• The pre-planned co-ordinates of this image, which is the desired output of the training pattern, is used to stimulate two related output units along the x and y output layer respectively The training data are automatically grabbed and segmented when one tracks with the eyes a cursor movement following a pre-designed zigzag path across or up and down the screen. A total of 2000 images are needed to train the system to achieve better performance.

Issues in gaze tracking system:

• Imprecise data, or the head may pan and tilt resulting in many eye images (relative to the viewing camera) corresponding to the same co-ordinates on the screen,

• Noisy images, mainly due to change of lighting. This is typical in an uncontrolled open plan office environment.

• Possibly infinitely large image set, in order to learn the variations of the images and make the system generalize better.

Conclusion

In this paper I have shown how such flexibility can be exploited within the context of an intelligent multi-modal interface. Soft computing techniques have been used at the interface for temporal reasoning, approximate query matching, learning action sequences and gaze tracking. It is important to note that soft computing has been used in conjunction with other AI- based systems performing dynamic scheduling, logic programming, speech recognition, and natural language understanding. Soft computing in combination with other AI techniques can make a significant contribution to human-centered computing in terms of development time, robustness, cost, and reliability.

Future work:

The following improvements can be made in the reasoner in the near future.

• The context can be extended to have a tree like structure such that the user is able to make reference to previously used contexts.

• The temporal reasoner can be used more extensively in conflict resolution.

• The grammar can be extended to include multiple format grammars.

• The dialogue system can be improved to become more user friendly.

• Different approaches can be used for resolving ambiguities such as competition between modalities or bidding.

Acknowledgment

I would like to acknowledge the SOFTCOMP’11 team for giving me an opportunity to present my paper.

References

[1] Shoemaker l., Nijtmans J., Camurri A., Lavagetto F., Morasso P., Benoit C., Guiard-Marigny A., Le Goff B., Robert-Ribes J., Adjoudani A., Defee I., Munch S., Hartung K. and Blauert J., 1997 ‘A Taxonomy of Multimodal Interaction in the Human Information Processing System’, Report of the ESPRIT PROJECT 8579 MIAMI, February.

[2] Tsui K.C., Azvine B., Djian D., Voudouris C., and Xu L.Q., 1998, Intelligent Multimodal Systems, BTTJ Journal, Vol.16, No. 3, July..

[3] Lesaint, D., C. Voudouris, N. Azarmi and B. Laithwaite 1997. Dynamic Workforce Management. UK IEE Colloquium on AI for Network Management Systems (Digest No. 1997/094), London, UK.

[4] Azvine, B., Azarmi, N. and Tsui, K.C. 1996. Soft computing - a tools for building intelligent systems, BT Technology Journal, vol. 14, no. 4, pp. 37 45, October.

[5] Bishop, C. 1995.Neural Network for Pattern Recognition. Oxford University Press.

[6] Negroponte, Nicholar 1995. Being Digital. Coronet Books.

[7] Nigay L., Coutaz J., 1995, A Generic Platform for Addressing the Multimodal Challenge, CHI ’95 Proceedings.