AI Activities at CSRL: An Overview

O. P. Baghel, U. V. Kadam, S. K. Kopparapu, B. Mathew, P. G. Passahna,
P. V. S. Rao*, A. C. Srivastava, P. Sundaram

Cognitive Systems Research Laboratory
Plot No 14, Sector 24, Opp. Sanpada Rly. Station
Vashi - Turbe, Navi Mumbai - 400 705, India.

November 26, 2001

Contents

*PVS.Rao@TataInfotech.Com

1 Introduction

Cognitive Systems Research Laboratory is the research wing of the Tata Infotech Limited. The CSR Laboratory is involved in the state of the art research activity primarily in the areas of speech, script, and natural language processing. The objective of the CSR Laboratory which was set up two years ago was to combine the word pool of knowledge and local creativity and innovation to conceptualize futuristic technologies, techniques, ideas and potential products, which would not only give Tata Infotech Limited a competitive edge but also enable the state of the art technology to reach the common man.

2 What are Cognitive Systems?

The words cognition and recognition are frequently used interchangeably common parlance. However, they do not signify the same thing. Recognition is the process of identifying 'something' as being the same as one that is already known. For example we may recognize something as being a Cow, a chair or as noise. To recognize something as a chair, you essentially compare it in some way with your mental pictures of a chair, a table, a stool and so on and then make an appropriate decision. This is a 'passive' process and implies no real intelligence.

In contrast, cognition is the process of 'knowing' or 'perceiving'. It implies understanding or interpreting the 'meaning' of a message or stimulus: for example one might interpret what one sees hears or experiences — say, as a smile, a painful cry or as a chase. It is quite easy to see that interpretation or realization is a more complex activity and requires thinking of a type which is not involved in the recognition examples cited above. Cognition requires intelligence. The human has been blessed with an inquisitive mind. His nature is to question the innate nature of the objects and phenomena around him. Life itself is a most Interesting subject to study, in all its aspects: emergence, survival, replication, adaptation — also evolution into more complex forms — eventually into intelligence as we find it in humans. Particularly fascinating is the evolution of intelligence and of the faculties of Speech, Vision and Language for humans: not because other faculties are less important, but because these are the main faculties of human cognition and because they account for most of the knowledge that humans acquire in their lives. A good way of understanding these faculties is to try to realize them in Computers. This also has the potential for major practical applications.

3 Activities

The CSR Lab aims to approach the problems of speech, script and Natural Language in a cognitive sense and not merely as problems of recognition or classification.

Artificial Intelligence (AI) can be loosely defined as the art of making intelligent computer programs that mimic human intelligence albeit in a very small way. The activities at CSR Laboratory revolve around spoken speech, language and written script (in English and other Indian Languages) in a way that enables us to understand and interpret the information embedded therein as a normal human would.

Our active research focus is in the area of computer understanding of natural languages. Our work centers around understanding questions in a formal way so as to be able to translate such understanding into implementable algorithms for various practical applications. This is

also supported by the Gartner forecast that a deeper penetration of natural language tools into e-service systems is expected.

4 Speech Research

Speech and language constitute the main faculties of human cognition. Speech and Language are the most spontaneous and primary avenues for exchange of information between humans. They account for over 70% of the information that a human acquires in his life. Vision accounts for most of the remaining 30%. Speech and language have undoubtedly played a major role in the emergence and enhancement of intelligence in humans, causing them to race ahead of all other aspects in this respect. For instance, intelligent thinking in any reasonable sense would not be possible without language of some sort. Implementing systems for visual pattern recognition and natural language systems in computers is intellectually very challenging. Moreover, they hold the potential for many practical applications.

Our work in the area of speech can be broadly classified into one of speech analysis and speech synthesis.

4.1 Recognition

CSR Laboratory is pursuing state-of-the-art research with the aim of developing a robust continuous speech recognition engine. The aimed engine would serve a wide variety of purposes. including biometric authentication in several applications. Towards this end, we have developed an Isolated Word Recognition engine, a Speaker Verification system which works with high accuracy and robustness. In the pipeline are a Keyword Spotting and a Connected Digit (number) recognition system.

4.1.1 Isolated Word Recognition

A fixed vocabulary Isolated Word Recognition engine works in three modes: speaker dependent, gender dependent-speaker independent and speaker independent. The recognition accuracy of the engine is over 90% for telephone speech (noisy) and is over 95% for direct (desktop; less noisy) speech. Some of the applications developed using this Isolated Word Recognition engine are

- **Name dialing** application for telephone speech: The engine has been trained to recognize names of people spoken through a telephone. An off-the-shelf telephone interface card (Intel Dialogic; Model No. VFX 40ESE) has been used to provide the audio interface to the speech system. A dialogue module suitable for data collection over the telephone has been specifically designed for this purpose.
- **Voice-activated web browsing** application for direct (desktop) speech: The recognizer has been interfaced with a rudimentary browser to enable voice-activated web browsing.

4.1.2 Speaker Verification System

A Speaker Verification System that enables verifying a person's identity based on his voice. Use of this biometric-based technique for verification on the basis of an individuals voice covers a wide spectrum of applications.

The system has been tested for speakers speaking over telephone as well as speakers speaking using a microphone. Speakers can choose any password of their choice (e.g. friend's telephone number, a person's name etc) and need to repeat their chosen password 5 times for enrolling with the system. The valid acceptance of this system is very high and valid rejection is also very high for the 15 enrolled speakers. Further tests with more speakers are being carried out. This accuracy is maintained even if the *impersonator* has access to the chosen password and uses it while attempting illegal access.

4.2 Synthesis

Text-to-Speech synthesizers have very good potential in the field of Computer and communications industry today. Voice Response Systems (VRS) with pre-recorded and spliced speech, currently being used in some common systems, are very restrictive. Text-to-speech synthesizer based systems, on the other hand, can render unconstrained speech. Their *native English* accent, however, is not acceptable to most Indian users particularly for Indian names.

The speech synthesis activity was aimed initially at the development of a text-to-speech formant synthesizer for Indian languages including Indian English. The Text-to-Speech synthesizer generates intelligible, reasonably natural sounding speech with an "Indian Accent".

Development is underway for a system with even more natural voice, incorporating the concepts of source modeling and customization to female and child voices. Implementation of prosody too is proposed to be developed.

This speech synthesizer can be integrated with many commercial applications: natural voice response systems over telephone for railways, airlines, the tourism industry, hotels etc. Other areas of application include voice Portals (Information Services), Text Readers for e-learning, email/news reading etc.

5 Natural Language Processing

A central part of intelligence is the ability to ask and answer questions. This requires language competence. Conversely, even use of language demands high intelligence. Thus, language competence itself is generally considered to be proof of intelligence. A person who does not have the ability to understand language or to articulate his thoughts can hardly be called intelligent. The mentally retarded find it very difficult to acquire language skills. Work in this area therefore has much intellectual appeal. There is also application potential.

NLP has already opened up new opportunities for providing better quality services to customers by providing them the information they require in an automated and online fashion, without requiring a user to know about any formal query language. Advanced text mining and grammar modeling algorithms have opened additional avenues NLP technology is highly promising with respect to the sharing of the huge information available on the web, by different users, who may belong to a particular linguistic community or require very specific information. Due to high complexity of language models there are still many challenging open ended problems. Despite the technology not having matured, it has been able to offer many satisfactory applications in different areas.

5.1 Online Answering System with Intelligent Sentience

Question answering systems are useful in a variety of applications and are of extreme importance, particularly in the context of e-commerce and web technology. Online Answering System with Intelligent Sentience, developed at the CSR Lab, is a question-answering answering system enabling man-machine interaction in common English. It is specifically oriented towards answering questions and providing information on a given topic.

It adopts a user-friendly conversational style in responding to user queries and in providing them with the information they ask for. The latest offering from Tata Infotech is a generic NLP engine that can be integrated with third party, domain specific, applications. The characteristics of the engine are briefly described below. The system is:

- meant for use online through a browser
- Specially oriented towards providing information on a given topic
- Tunable to various domains
- Senses the intent of the user even if the question is poorly framed
- Handles typos in user queries
- Chooses the best answer from the alternatives available in the knowledge base.
- Presents answers with a prelude that gives a personal touch

5.2 Pidgin - Sentence Generation Tool

An interactive tool that enables a user to write meaningful sentences in a language that he is barely familiar with. The idea is extendable such that a user can generate text of arbitrary complexity in any other language that he does not even know (say English and Japanese, French, Hindi etc).

5.3 Automated Tender Forwarding

A system for extracting information from text documents, tabulating this information into semi-structured format and then routing it to the concerned individuals, based on a set of criteria they specify.

5.4 Question Understanding

This is an ongoing activity and we are in the process of trying to formalize our understanding so that it can be translated into a machine implementable scheme.

6 Script

Our interest in script is primarily due to its similarity with speech in the sense that it is a signal generated by the human and is meant to convey complex information effectively and efficiently.

CSR Laboratory's activity spreads not only into the analysis and recognition of handwriting cursive script but also into understanding and synthesizing script.

Our work primarily concentrates on finding a set of parameter or shape vectors, which are simple to extract but at the same time are powerful enough to achieve in variance despite the inherent variabilities in the written script.

6.1 Script Synthesis

Script synthesis involves essentially a mechanism of being able to extract salient features, which are smaller in number than the original data set so that the feature set could be used to replicate the script without any significant loss of spatial information.

Script synthesis not only serves in compressing data (typically the parameter set constitutes a small fraction of the total data representing the original signal) but also is useful for the purpose of script recognition and connected handwritten script generation from individual characters.

The other aspect of script synthesis is the idea of being able to generate a connected handwritten script from individual characters written in isolation. In this case, the so called transition strokes for concatenation of individual character shapes are synthesized using the concept of *minimal effort* to generate connected script. Typically, this method can be used to generate a personalized handwritten document from its roman equivalent individual alphabets of a person are taken and the transitions are synthesized to generate connected script words which will replicate the way it would have been written by the individual.

6.2 Recognition

We have developed an approach for word-based on-line and off-line recognition for cursive script composed of lower-case English letters. The system uses simple and easily extractable shape features such as direction of movement and curvature and relative locations of regions where these suffer discontinuities. Our approach is based on an empirical and robust parametric model for the hand system.

It addresses the problem of recognizing handwriting (lower case English letters) from data collected with a digitizer, which provides (digitized pen trajectory) temporal information. The basic advantage of the system is that it does not involve any training what so ever since it is based only on the broad shape information which is extracted and quantized to make the system robust. This work may lead to a product for a hand held script input device for computers and signature verification for ATM machines at banks.

7 Conclusions

Research in frontier areas of speech, script and natural language processing is an ongoing activity at the Cognitive Systems Research Laboratory. The choice of the research areas is primarily because of the expertise available within the laboratory and also because the research is challenging in the sense that it provides us a platform to research newer dimensions in understanding human cognition systems. The basic aim of the laboratory is to utilize the pool of knowledge available world wide and to contribute to this pool in a significant way with the aim of developing technology and techniques that would be relevant for Tata Infotech Limited and give it a technological edge.