A Review of Machine Translation Using Various Soft-Computing Tools

Manish Rand, Mohammad Atique
Thakur College of Engineering & Technology,
Mumbai, India
manishrana@live.com

ABSTRACT: *In this paper, there is description about various techniques used for machine translation. It simply states that how natural language processing had played a major role. This paper shows a review of techniques used for machine translation. Various techniques and method people have discuss to reveal the work. The techniques have used American Sign Language (ASL), Optimization method, Automatic speech recognition (ASR) which have show a significant result for translation. Machine Translation has made great strides with the development of digital signal processing hardware and software Fuzzy sets. The process of Fuzzification provides a mechanism by which real-valued features is also used for machine translation. Another method Approximate Reasoning is the process or processes by which a possible imprecise conclusion is deduced from a collection of imprecise premises. Maximum mutual information (MMI) and a novel "error weighted" training technique exemplar-based sparse representations for noise robust automatic speech recognition has also been developed. Going through all these research and methods to develop a machine translation is not enough to conclude a fruitful result.*

Keywords: Fuzzification, Fuzzy Logic, Optimization,. Exemplar-based, Machine Translation, Machine Translation, American Sign Language (ASL), Automatic Speech Recognition (ASR), Maximum Mutual Information (MMI) etc

Received: 19 November 2014, Revised 24 December 2014, Accepted 29 December 2014

1. Introduction

Starting with an example of machine translation, English to Hindi Translation provides the most convenient access to online translation service powered by various machine translation engines. English to Hindi Translation tool includes online translation service, English text-to-speech service, English spell checking tool, on-screen keyboard for major languages, back translation, email client and much more. The most convenient translation environment is ever created.

Current machine translation software often allow for customization by domain or profession (such as weather reports), improving output by limiting the scope of allowable substitutions. This technique is particularly effective in domains where formal or formulaic language is used. It follows that machine translation of government and legal documents more readily produces usable output than conversation or less standardized text.

Improved output quality can also be achieved by human intervention: for example, some systems are able to translate more accurately if the user has unambiguously identified which words in the text are proper names. With the assistance of these techniques, MT has proven useful as a tool to assist human translators and, in a very limited number of cases, can even produce output that can be used as is (e.g., weather reports).

The progress and potential of machine translation have been debated much through its history. Since the 1950s, a number of scholars have questioned the possibility of achieving fully automatic machine translation of high quality. Some critics claim that there are in-principle obstacles to automat zing the translation process. Thus discussing various work carried out in last few year in case of machine translation.

2. Literature Survey

In the literature survey, there are many papers that show with the graphs the works and techniques used for machine transaltion. All the papers are mentioned below with their work.

2.1 Toward An Example - based Machine Translation From Written Text to ASL Using Virtual Agent Animation

In this paper authors discuss about Modern computational linguistic software [1] that cannot produce important result for sign language translation. The author uses some research methodology to find that majority of automatic sign language system for translation and ignore many other aspects they generate; therefore the interpretation lost the truth information meaning. This problem is due to sign language consideration as a derivative but it is a complete language with its own unique grammar as described in it related to other paper referred. This grammar is related to semantic-cognitive models of spatially, time, action and facial expression to represent complex information to make sign interpretation more efficiently, smooth, expressive and natural-looking human gestures. Author says all this aspects give them useful insights into the design principles that have evolved in natural communication between people. But not significant result is shown in the steps of machine translation.

2.2 Optimization Algorithms and Applications for Speech and Language Processing

This authors talk about Optimization techniques [2] that have been used for many years in the formulation and solution of computational problems arising in speech and language processing. Such techniques are found in the Baum-Welch, extended Baum-Welch (EBW), Rprop, and GIS algorithms, for example. Additionally, the use of regularization terms has been seen in other applications of sparse optimization. This paper outlines a range of problems in which optimization formulations and algorithms play a role, giving some additional details on certain application problems in machine translation, speaker/language recognition, and automatic speech recognition. Several approhyaches developed in the speech and language processing communities are described in a way that makes them more recognizable as optimization procedures

Literature Description:

According to author, the purpose of this paper is to review a number of key application areas from the speech and language processing literature, focusing in particular on the way in which the important data processing problems in these areas are formulated and solved. To explain the formulations and algorithms in a common optimization framework, identifying clearly the variables, objectives, and constraints. It also explores the relationship between the algorithms derived in the application communities and the model-based framework that is the basis of many important optimization algorithms. By placing past work in these areas more firmly in the context of optimization, and thus more easily comprehensible to those outside the communities in which this work was originally performed, author hopes to make it easier to identify possible enhancements to the algorithms, and thus to stimulate new research at the intersection of optimization and speech/language processing.

2.3 Feature Selection Algorithm for Automatic Speech Recognition Based On Fuzzy Logic

In this paper authors describe about Automatic speech recognition [3] (ASR) for the development of digital signal processing hardware and software in case of speech recognition. Author focus on that inspite of automatic speech recognition, machines cannot match the performance of their human counterparts in terms of accuracy and speed, especially in case of speaker independent speech recognition. Many papers present the viability of Mel Frequency Cepstral coefficient Algorithm to extract features and Fuzzy Inference System model for feature selection, by reducing the dimensionality of the extracted features. There is an increasing need for a new Feature selection method, to increase the processing rate and recognition accuracy of the classifier, by selecting the discriminative features. Hence author concludes a Fuzzy Inference system model is used selecting the optimal features from speech vectors which are extracted using MFCC. The author feature concludes that work has been done on MATLAB 13a and experimental results show that system is able to reduce word error rate at sufficiently high accuracy. But it does fail to give efficiency in machine translation.

Literature Description

Author explains that the main goal of speech recognition area is to develop techniques and systems for speech input to machine. Speech is the primary means of communication between humans. Based on major advances in statistical modeling of speech, automatic speech recognition systems today find widespread application in tasks that require human machine interface, such as automatic call processing in telephone networks, and query based information systems that provide updated travel information, stock price quotations, weather reports, Data entry, voice dictation, access to information: travel, banking, Commands, Avionics, Automobile portal, speech transcription, Handicapped people (blind people) supermarket, railway reservations etc. Speech recognition technology was increasingly used within telephone networks to automate as well as to enhance the operator services. Thus speech recognition plays a major role in most of the applications. The basic model of speech recognition is shown in the figure 1.

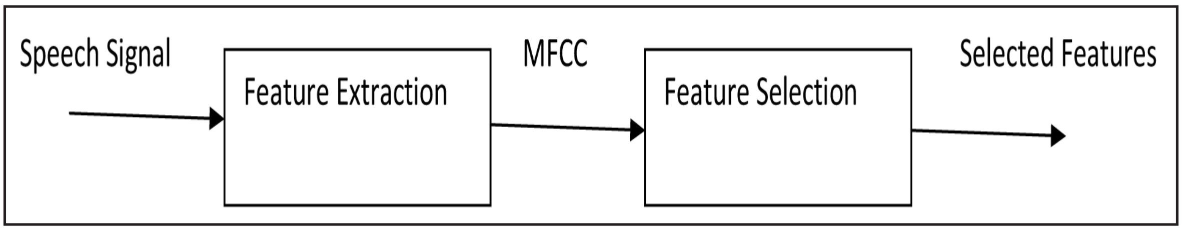

Figure 1. Basic Model of Speech Recognition

Feature Selection

Author further describes that the feature selection can be viewed as one of the most fundamental problems in the field of machine learning. The main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. In real world problems, feature selection is a must due to the abundance of noisy, irrelevant or misleading features. By removing these factors, learning from data techniques can benefit greatly. Fuzzy sets and the process of Fuzzification provide a mechanism by which real-valued features can be effectively managed . By allowing values to belong to more than one label, with various degrees of membership, the vagueness present in data can be modeled. The feature selection phase is performed by a fuzzy inference system based on the set of rules obtained from the Mel frequency coefficients. The extracted 39 coefficients are used by the fuzzy inference system to generate Gaussian membership functions.

Fuzzy Inference Systems

Fuzzy inference systems are also known as fuzzy rule-based systems. Basically, a fuzzy inference system is composed of four functional blocks is shown in figure 2

1. A Knowledge base, containing a number of fuzzy rules and the database, which defines the membership functions used in the fuzzy rules.

2. An Inference engine, which performs the inference operations on the rules.

3. A Fuzzification interface, which transforms the crisp inputs into degrees of match with linguistic values.

4. A Defuzzification interface, which transforms the fuzzy results of the inference into a crisp output.

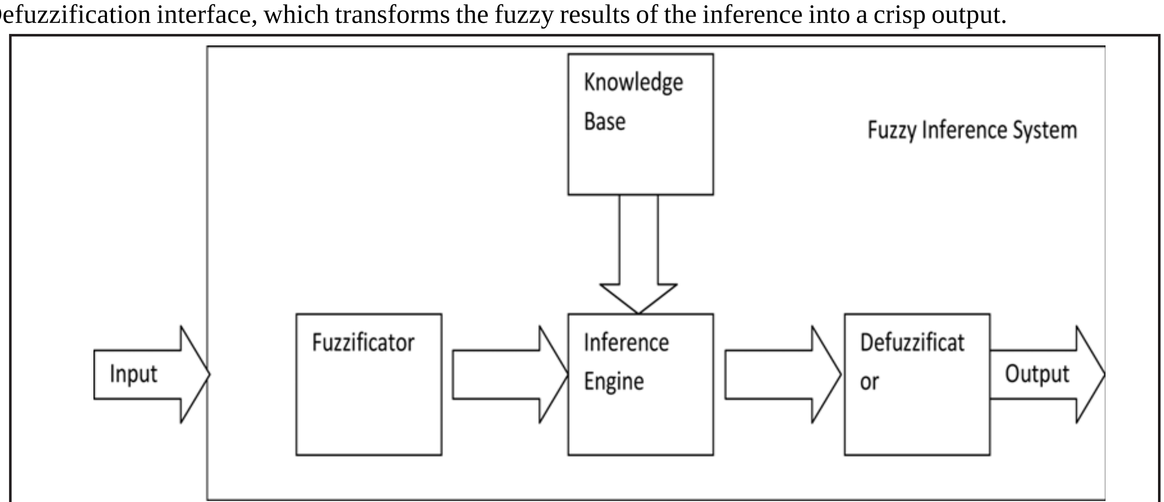

Figure 2. Fuzzy Inference System

2.4 Fuzzy Logic and Approximate Reasoning: An Overview

This paper authors describe about Approximate Reasoning [4] which is the process or processes by which a possible imprecise conclusion is deduced from a collection of imprecise premises. Fuzzy logic plays the major role in approximate reasoning. It has the ability to deal with different types of uncertainty.

An overview of the different aspects of the theory of approximate reasoning has been provided here based on the existing literature. Suitable illustrations are included, whenever necessary, to make the concept clear. Some of the implementations of the theory to real life problems have been mentioned. Finally, a linguistic recognition system based on approximate reasoning has been described along with its implementation in speed1 recognition problem.

Linguistic Variable

Authors have used the basic concept in fuzzy logic that plays a key role in approximate reasoning is a linguistic variable, which in early seventies was called a variable of higher order rather than a fuzzy variable. A linguistic variable, as its name suggests, is a variable whose values are not numbers but words or sentences in a natural language. For example, height is a linguistic variable if its values are linguistic rather than numerical, ie, short, tall, very short, not short, not very tall, quite tall, not very short and not very tall etc, rather than 100, 110, 120, ... (In cm). In general, the values of a linguistic variable can be generated from a primary term (eg, "*tall*"), its antonyms (eg, "*short*"), a collection of modifiers (eg, "*not*", "*very*", "more or less", "quite", "*not very*" etc) and the connectors ("*and*" and "*or*")

Definition

A linguistic variable is characterized by a quintuple (X, T (X), U, G, M) in which X is the name of the variable; (X) is the term set of X; U is a universe of discourse; G j s a syntactic rule which generates the terms in r (X); and M is a semantic rule which associates with each linguistic value X its meaning where M(X) denotes a fuzzy subset of U. For a particular X, the name genera ted by G, is called a term.

Example

Suppose X is a linguistic variable with the label "*Height*" with U = [0,250}. Terms of this linguistic variable, which are fuzzy sets, could be called "*tall*", "*short*", "*very tall*" and so on.

2.5 Discriminative Estimation of Subspace Constrained Gaussian Mixture Models for Speech Recognition

In this paper authors make a study discriminative training of acoustic models for speech recognition [5] under two criteria: maximum mutual information (MMI) and a novel "error weighted" training technique. The author presents a proof that the standard MMI training technique is valid for a very general class of acoustic models with any kind of parameter tying. We report experimental results for subspace constrained Gaussian mixture models (SCGMMs), where the exponential model weights of all Gaussians are required to belong to a common "*tied*" subspace, as well as for Subspace Precision and Mean (SPAM) models which impose separate subspace constraints on the precision matrices (i.e. inverse covariance matrices) and means. It has been shown previously that SCGMMs and SPAM models generalize and yield significant error rate improvements over previously considered model classes such as diagonal models, models with semi-tied covariance's, and EMLLT (extended maximum likelihood linear transformation) models. We show here that MMI and error weighted training each individually result in over 20% relative reduction in word error rate on a digit task over maximum likelihood (ML) training. Author also shows that a gain of as much as 28% relative can be achieved by combining these two discriminative estimation techniques.

2.6 Ensemble Deep Learning for Speech Recognition

This paper authors describe about Deep learning [6] systems that have dramatically improved the accuracy of speech recognition, and various deep architectures and learning methods have been developed with distinct strengths and weaknesses in recent years. How can ensemble learning be applied to these varying deep learning systems to achieve greater recognition accuracy is the focus of this paper. Author develop and report linear and log-linear stacking methods for ensemble learning with applications specifically to speech-class posterior probabilities as computed by the convolution, recurrent, and fully-connected deep neural networks. Convex optimization problems are formulated and solved, with analytical formulas derived for training the ensemble-learning parameters. Experimental results demonstrate a significant increase in phone recognition accuracy after stacking the deep learning subsystems that use different mechanisms for computing high-level, hierarchical features.

from the raw acoustic signals in speech.

Literature:
Linear Ensemble

To simplify the stacking procedure for ensemble-learning, Author perform the linear combination of the original speech-class posterior probabilities produced by deep learning subsystems at the frame level here. A set of parameters in the form of full matrices are associated with the linear combination, which are learned using the training data consisting of the frame-level posterior probabilities of the different subsystems and of the corresponding frame-level target values of speech classes.

2.7 Exemplar- Based Sparse Representations for Noise Robust Automatic Speech Recognition

This paper author proposes to use exemplar-based [7] sparse representations for noise robust automatic speech recognition. First, Author describe how speech can be modeled as a linear combination of a small number of exemplars from a large speech exemplar dictionary. The exemplars are time-frequency patches of real speech, each spanning multiple time frames. Author then propose to model speech corrupted by additive noise as a linear combination of noise and speech exemplars, and authors derive an algorithm for recovering this sparse linear combination of exemplars from the observed noisy speech. Authors describe how the framework can be used for doing hybrid exemplar-based/HMM recognition by using the exemplar-activations together with the phonetic information associated with the exemplars. As an alternative to hybrid recognition, the framework also allows us to take a source separation approach which enables exemplar-based feature enhancement as well as missing data mask estimation. We evaluate the performance of these exemplar based methods in connected digit recognition on the AURORA-2 database. Our results show that the hybrid system performed substantially better than source separation or missing data mask estimation at lower SNRs, achieving up to 57.1% accuracy at SNR= -5 dB. Although not as effective as two baseline recognizers at higher SNRs, the novel approach offers a promising direction of future research on exemplar-based ASR.

Literature:
Authors use of a speech dictionary containing exemplars as atoms has several advantages. First, the dictionary is relatively easy to construct by extraction of speech segments from a speech database. Second, it becomes computationally efficient to construct dictionaries with high-dimensional atoms that contain several frames of time context, which makes confusion between noise and speech atoms in less likely. Third, the dictionary can allow for very sparse representations if an observed speech segment closely resembles speech contained in the dictionary [22]. Finally, the use of exemplars makes the mapping from atoms to speech classes straightforward: Each time-frame in the speech exemplars is directly labeled with an HMM-state label, obtained by means of a forced alignment of the transcription on the training database using a conventional HMM-based recognizer.

Sparse Representation of Noisy Speech

Speech signals are represented by their spectra-temporal distribution of acoustic energy, a spectrogram. The exemplar based approaches proposed in this paper operate in the Mel scale magnitude spectrogram domain, with the term magnitude referring to the square root of energy in a time-frequency element. The cepstral features used in conventional ASR systems are based on a (mostly logarithmic) compression of the magnitude values followed by a decor relating cosine transform. In our framework, however, we use the magnitude values directly to simplify the additively of speech and noise.

2.8 Discriminative Learning in Sequential Pattern Recognition

This article author is motivated [8] by the striking success of the MMI, MCE, and MPE/MWE-based discriminative criteria in speech recognition. Yet in the past, there was a lack of common understanding of the interrelation among these techniques, despite the relatively long history of MMI ,MCE and MPE/MWE . Due to the complexity of these techniques and the lack of a common underlying theoretical theme and structure, disparate discriminative learning procedures were developed and parameter optimization has become a major issue. The main goal of this article is to provide an underlying foundation for MMI, MCE, and MPE/MWE at the objective function level to facilitate the development of new parameter optimization techniques and to incorporate other pattern recognition concepts, e.g., discriminative margins , into the current discriminative learning paradigm.

2.9 Example-based Machine Translation Based on Syntactic Transfer with Statistical Models

Author of this paper presents example-based machine translation (MT) based on syntactic transfer [9], which selects the best translation by using models of statistical machine translation. Example-based MT sometimes generates invalid translations because it selects similar examples to the input sentence based only on source language similarity. The method proposed in this paper selects the best translation by using a language model and a translation model in the same manner as statistical MT, and it can improve MT quality over that of 'pure' example-based MT. A feature of this method is that the statistical models are applied after word re-ordering is achieved by syntactic transfer. This implies that MT quality is maintained even when we only apply a lexicon model as the translation model. In addition, translation speed is improved by bottom-up generation, which utilizes the tree structure that is output from the syntactic.

Literature Survey:

In this paper author presents an example-based MT method based on syntactic transfer, which selects the best translation by using models of statistical MT. This method is roughly structured using two modules (Figure 2). One is an example-based syntactic transfer module. This module constructs.

Figure 3: Structure of Proposed Method tree structures of the target language by parsing and mapping the input sentence while referring to transfer rules. The other is a statistical generation module, which selects the best word sequence of the target language in the same manner as statistical MT. Therefore, this method is sequentially combined example-based and statistical MT. The proposed method has the following advantages.

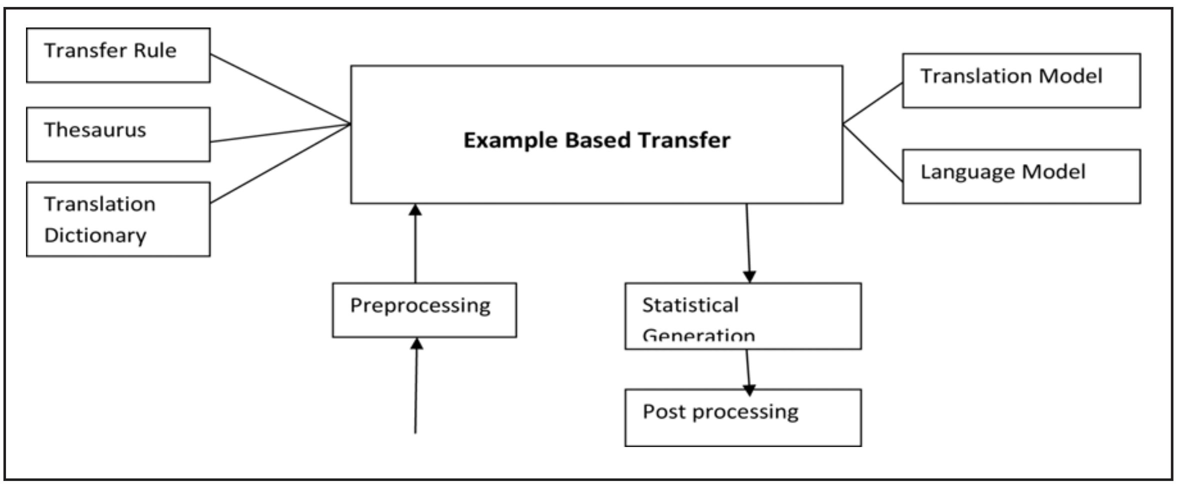

Figure 3. Structure of Propose Method

3. Result and Comparison

S.No	Papers Technique	Method	Sequence	Alignment	Result
1.	Bilingual Machine Transliteration (BMT)	×	√	×	√
2.	extended Baum-Welch (EBW)	×	√	×	√
3.	Knowledge Based Machine Translation System (KMTS)	√	×	√	×
4.	Statistical Machine Translation (SMT)	×	√	×	√
5.	Example based Machine Translation (EBMT)	×	×	√	×
6.	Machine translation (MT)	√	×	×	√
7.	GIS Algorithm	√	√	×	√

Table 1. Comparison of Various Techniques used for Machine Translation

S.No.	Parameter	RBMT	SMT	EBMT	SVM
1	Consistency	High	Low	Medium	Medium
2	Predictable Quality	Good	Similar	Very well	Very Good
3	Out of Domain Quality	Medium	Low	High	Medium
4	Use of Grammar	Yes	No	No	Yes
5	Robust	Yes	No	Yes	No
6	Fluency	Less	Medium	High	Medium
7	Performance	Good	Medium	Good	Medium

Table 2. Comparison of various Machine Translation schemes

Table 1 shows the comparison used in various technical work proposed and implemented in different papers and **Table 2** illustrates the comparison of three machine translation techniques, Rule-Based Machine Translation (RBMT), Statistical Machine Translation (SMT), Example-Based Machine Translation (EBMT) and support vector machine(SVM) on the basis of various parameters such as Consistency, predictable quality, Quality of out of domain translation, Use of grammar, robustness, Fluency and performance.

Acknowledgments

In this paper various paper techniques are discussed which had been derived by different authors .Thanking all the author to describe and used there methods for machine translation for betterment of machine learning.

References

[1] Boulares, Mehrez., Jemni, Mohamed. (2012). Research Lab. UTIC, University of Tunis, 5, Avenue Taha Hussein, B. P. : 56, Bab Menara, 1008 Tunis, Tunisia, Toward an example-based machine translation from written text to ASL using virtual agent animation, *IJCSI International Journal of Computer Science Issues*, 9, (1) 1 January 2012

[2] Stephen, J., Wright., Kanevsky, Dimitri. (2013). Senior Member, IEEE, LiDeng, Fellow, IEEE,, Xiaodong He, Senior Member, IEEE, Georg Heigold, Member, IEEE, and Haizhou Li, Senior Member, IEEE, Optimization Algorithms and Applications for Speech and Language Processing, *IEEE Transactions on Audio, Speech and Language Processing*, 21 (11) 2231-2243

[3] Nereveettil, Catherine J., Kalamani, M., Valarmathy, S. (2014). Feature Selection Algorithm for Automatic Speech Recognition Based On Fuzzy Logic, International Journal of Advanced Research in Electrical, *Electronics and Instrumentation Engineering*(An ISO 3297: 2007 Certified Organization) 3(1) January 2014.

[4] Pal, Sankar, K., Mandal, Deba Prasad (1991). Electronics and Communication Science Unit, Indian Statistical Institute, calcutta 700 035, India Fuzzy Logic and Approximate Reasoning: An Overview, Reprinted from the *Journal of the Institution of Electronics and Telecommunication Engineers*, 37(5&6) 548-560

[5] Axelrod, Scott ., Goel, Vaibhava., Gopinath, Ramesh., Olsen, Peder, Visweswariah, Karthik (2004). Discriminative Estimation of Subspace Constrained Gaussian Mixture Models for Speech Recognition, *IEEE Trans on Speech and AUdio Processing*, March 10, 1- 18

[6] Deng, Li., Platt, John C. Ensemble Deep Learning for Speech Recognition, Microsoft Research, One Microsoft Way, Redmond, WA, USA. 568-5674

[7] Gemmeke, Jort F. (2011). Student-Member, IEEE, Tuomas Virtanen, Antti Hurmalainen Exemplar-based sparse representations for noise robust automatic speech recognition, (c) 2011 IEEE. 1-14.

[8] He, Xiao dong., Li Deng., Chou, Wu. (2008). Discriminative Learning in Sequential Pattern Recognition, *IEEE Signal Processing Magazine, p.*19- 26

[9] Imamura, Kenji., Okuma, Hideo., Watanabe, Taro., Eiichiro Sumita. Example-based Machine Translation Based on Syntactic Transfer with Statistical Models, ATR Spoken Language Translation Research Laboratories 2-2-2 Hikaridai, Keihanna Science City, Kyoto, 619-0288, Japan, p. 634-641.