

Identification of Question and Non-Question Segments in Arabic Monologues Using Prosodic Features: Novel Type-2 Fuzzy Logic and Sensitivity-Based Linear Learning Approaches

Sunday Olusanya Olatunji*, Lahouari Cheded, Wasfi G. Al-Khatib, Omair Khan

College of Computer Science and Engineering, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.
Email: oluolatanji.aadam@gmail.com, cheded@kfupm.edu.sa, wasfi@kfupm.edu.sa, omikhan@gmail.com

Received November 8th, 2012; revised January 5th, 2013; accepted January 14th, 2013

Copyright © 2013 Sunday Olusanya Olatunji *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

In this paper, we extend our previous study of addressing the important problem of automatically identifying question and non-question segments in Arabic monologues using prosodic features. We propose here two novel classification approaches to this problem: one based on the use of the powerful type-2 fuzzy logic systems (type-2 FLS) and the other on the use of the discriminative sensitivity-based linear learning method (SBLLM). The use of prosodic features has been used in a plethora of practical applications, including speech-related applications, such as speaker and word recognition, emotion and accent identification, topic and sentence segmentation, and text-to-speech applications. In this paper, we continue to specifically focus on the Arabic language, as other languages have received a lot of attention in this regard. Moreover, we aim to improve the performance of our previously-used techniques, of which the support vector machine (SVM) method was the best performing, by applying the two above-mentioned powerful classification approaches. The recorded continuous speech is first segmented into sentences using both energy and time duration parameters. The prosodic features are then extracted from each sentence and fed into each of the two proposed classifiers so as to classify each sentence as a Question or a Non-Question sentence. Our extensive simulation work, based on a moderately-sized database, showed the two proposed classifiers outperform SVM in all of the experiments carried out, with the type-2 FLS classifier consistently exhibiting the best performance, because of its ability to handle all forms of uncertainties.

Keywords: Arabic Monologues; Prosodic Features; Type-2 Fuzzy Logic Systems; Sensitivity Based Linear Learning Method; Support Vector Machines

1. Introduction

There has been a huge increase in the amount of data generated and stored as computers and Internet are increasingly becoming part of our everyday life. This huge information exists in various formats: text, audio and video formats. With the availability of broader bandwidths in internet communication, there has been an increase in audio and video content on the Internet, in addition to text and image data that people were earlier used to. Online video and audio sites such as yahoo video, YouTube, Flickr, Google video, etc. are among the most visited sites on the Internet. Audio and video contents are widely shared through file-sharing peer-to-peer networks.

Multimedia content now constitutes the bulk of the Internet traffic in the form of IP-telephony, video and audio conferencing, Internet radio stations, music stores, lecture sites etc.

Audio data forms an important part of this multimedia data present and transmitted through the Internet. It includes online lectures, music, radio programs, podcasts, news, Text-To-Speech (TTS) systems that translate textual websites into audio for visually impaired people etc. This content is presented in downloadable as well as in streaming format, and is produced in a variety of languages. Most of it is in English and in other Western languages because of the digital divide between the West and the Rest. Prominent Non-Western languages on the

*Corresponding author.

Web include Japanese, Chinese, Korean, Turkish and Arabic. Arabic is the tenth most widely-used language on the Internet [1]. In the Arabic language, huge repositories of lectures on the Islamic faith are available [2,3]. An interesting problem in this regard is the semantic indexing of these lectures. This can be achieved by transcribing all the lectures manually or automatically through speech recognition techniques. However, neither solution is viable at this point. Note here that manual transcription is a very laborious process requiring manpower and time, and consequently comes with a high financial cost. Unlike speech recognition in other Western languages, Arabic speech recognition has not matured enough to achieve high accuracy rates on such unrestricted domain.

Therefore, one needs to look for other reasonable semantic content that can be automatically extracted. One such semantic content is that of questions posed within these lectures, whether by the speaker or the audience. In many such lectures, a question-answer session usually follows the main speech. These questions and answers form a sizable and very useful knowledge-base on various issues, which users may utilize. For example, one can look for the opinion of different scholars on a certain contemporary issue. Normally, the lectures are monologues, *i.e.* questions and answers are spoken by the same person. Therefore, the problem can be approached from a prosodic point of view, where the intonation of the voice is used for cues. Prosody is widely used in many speech-related applications. Researchers have used prosodic features for dialog act detection [4,5]; topic segmentation [6]; sentence boundary detection [7]; emotion detection [8] etc. Generally, people employ prosody to identify question segments in the speech in order to achieve high success rates. In the cited approaches, researchers have used various types of classifiers such as decision trees, support vector machines (SVM), artificial neural networks, Boosting algorithms like AdaBoost etc. Some authors, such as Lee and Narayanan [8], have improved the recognition accuracy by fusing the outputs from multiple classifiers into a single and more accurate output.

In this paper, we proposed two novel approaches, based on type-2 fuzzy logic systems (type-2 FLS) and the sensitivity-based linear learning method (SBLLM), to the identification of question and non-question segments in a monologue, using on prosodic features. The main motivation behind the adoption of these novel approaches is that they can handle uncertainties in a much better way than did previously-used techniques. This point is all the more important in our case since our experiments have been conducted on only a small-to-moderately-sized database, where achievable recognition rates can be adversely affected by large uncertainties caused by the

paucity of available data. It is to be pointed out at this juncture that the building of a large database is time-consuming and costly and has a major impediment in our experiments.

Type-2 fuzzy logic inference systems have been recently proposed as new intelligence frameworks for both prediction and classification to handle all forms of uncertainties, see [9-12] for more details. Type-2 fuzzy logic is powerful in handling uncertainties, including uncertainties in measurements and data used to calibrate the parameters. It was introduced as an extension of type-1 fuzzy logic to compensate for the inadequacy of type-1 fuzzy inference system. It has since featured in a wide range of medical, business and engineering applications, often with promising and consistent results [13,14]. Thus, the significance of type-2 fuzzy logic-based identification model, which is one of the models hereby proposed, cannot be overemphasized.

Sensitivity-based linear learning method (SBLLM) has been introduced as a learning technique for two-layer feed forward neural networks based on sensitivity analysis that uses a linear training algorithm for each of the two layers [15]. This algorithm tends to provide a good generalization performance at extremely fast learning speeds, while giving the sensitivities of the sum of squared errors with respect to the input and output data at no extra computational cost. It is very stable in performance as its learning curve stabilizes soon, and behaves homogeneously not only when considering just the end of the learning process, but also during the whole process, in such a way that very similar learning curves can be obtained for all iterations of different experiments [15, 16].

The rest of this paper is organized as follows. Prosodic features and the feature extraction process are presented in Section 2. Section 3 presents the proposed computational intelligence frameworks based on the two proposed novel approaches of type-2 FLS and SBLLM. Section 4 provides empirical studies and experiments, results, discussion, and comparative studies. The conclusion and further work are presented in Section 5.

2. Prosodic Features

Prosodic features can be divided into acoustic prosodic features and linguistic prosodic features [17]. Acoustic prosodic features are signal-based features that usually span segments larger than phonemes such as words, syllables, turn etc. They are usually extracted from the audio signal directly. Linguistic-prosodic feature are based on lexical information that is obtained from the word or phone transcript of a conversation, such as the marking of word-final syllables and features based on syntax or semantics of an utterance. In this study, we only use

acoustic prosodic features as no word information is assumed to be present and/or extracted. The following includes a listing of the major features used. More information on the features is elaborated on in [17].

- **Pitch Features:** The fundamental frequency or $F0$ feature is one of the most widely used pitch features. $F0$ values are normalized over their mean and standard deviation for each speaker across the whole conversation side. Change in pitch contour is another important set of pitch features that can be important in determining the utterance type.
- **Energy Features:** One feature that has been commonly used in this research is that of spectral balance. Spectral balance measures the energy distribution over the frequency spectrum. The overall intensity of sentence final syllables and standard RMS energy features were also used in this study.
- **Duration Features:** The term duration refers to the length of a particular constituent of speech. It has been noted that the utterance's final syllable is usually longer in question intonation than in statement intonation. Some of the duration features considered here include the duration of the final syllable, the average duration of other syllables, and the length of the whole utterance. Another feature that is helpful for sentence boundary detection is the slowing down toward the end of the units.
- **Pause Features:** Pause features help in dividing an utterance into sentences. The signal energy is most often used to calculate the pause features, most of which are extracted at word boundaries. Since we do not have any information about word and phones, some minimum energy threshold was empirically chosen so that any speech below that threshold would be considered as a pause. Pause features are often normalized according to the speaking rate of a particular speaker. If the speaking rate is high, then pauses will be of lesser duration than those in the low speaking-rate speech. We computed both raw and normalized features.
- **Speaking Rate Features:** Speaking rate is commonly defined as either the word-per-minute rate or the syllable-per-second rate. Since we do not have any information about either word or syllable, we considered the average length of voiced portion of speech in our work.

2.1. The Data Corpus

To the best of our knowledge, no corpus of questions and non-question segments has been developed for Arabic. We acquired 15 Arabic audio lectures by three male speakers with a sampling rate of 8 KHz and having a mono channel. We had 9294 sentences in total, out of

which only 514 could be classified as questions. In order to keep an equal prior probability of question and non-question classes during the training stage, we had to downsample the non-questions to 514 for a total of 1028 sentences. These sentences were obtained from the audio files by segmenting them using a pause as a criterion for separation. Then, features were calculated from those sentences. The process of feature extraction is further explained in [17].

One hundred and forty four features were extracted from each of these 1028 sentences. These 144 features were then reduced to 123 after removing basic pattern features. Basic pattern features had information in the nominal form and that information was then used to calculate derived pattern features. Details of the number of features in each feature set are given in **Table 1**.

3. Methodology and Operations of the Proposed Approaches

Briefly, the main theoretical points underlying the two proposed approaches of type-2 fuzzy logic systems (Type-2 FLS) and sensitivity based linear learning method (SBLLM), are presented next.

3.1. The Intelligent Framework Based on Type-2 Fuzzy Logic Systems (Type-2 FLS)

Type-2 adaptive fuzzy inference systems (ANFIS) is an adaptive network that learns the membership functions and fuzzy rules, from data, in a fuzzy system based on type-2 fuzzy sets, see [9,18] for details. Type-2 fuzzy sets are fuzzy sets whose grades of membership are themselves fuzzy. They are intuitively appealing because grades of membership can never be obtained precisely in practical situations [19]. The concept of a type-2 fuzzy set was introduced by Zadeh [20] as an extension of the concept of type-1 fuzzy set. A type-2 fuzzy set is characterized by a fuzzy membership function, that is, the membership grade for each element of this set is a fuzzy set in $[0, 1]$, unlike a type-1 set where the membership grade is a crisp number in $[0, 1]$. Such sets can be

Table 1. Feature sets and their dimensions.

Feature Set	No. of Features
Full	123
$F0$ -only	55
Energy-only	38
Patterns-only	31
$F0$ -Energy	92
$F0$ -Pattern	77
Energy-pattern	64

used in situations where there is uncertainty about the membership grades themselves, for example, an uncertainty in the shape of the membership function or in some of its parameters. Consider the transition from ordinary sets to fuzzy sets; when we cannot determine the membership of an element in a set as 0 or 1; we use fuzzy sets of type-1. Similarly, when the situation is very fuzzy that we have difficulty in determining the membership grade as a crisp number in $[0, 1]$, we use fuzzy sets of type-2.

In general, “a fuzzy set is of type n , $n = 2, 3, \dots$ if its membership function (MF) ranges over fuzzy sets of type $(n-1)$ ” [20]. Imagine blurring the type-1 membership function depicted in **Figure 1** by shifting the points either to the left or to the right and not necessarily by the same amounts, as in **Figure 2**. Then, at a specific value of x , say x' , there is no longer a single value for the membership function (μ'); instead, the membership function takes on values wherever the vertical line intersects the blur. Those values need not all to be weighted the same; hence, we can assign an amplitude distribution to all of

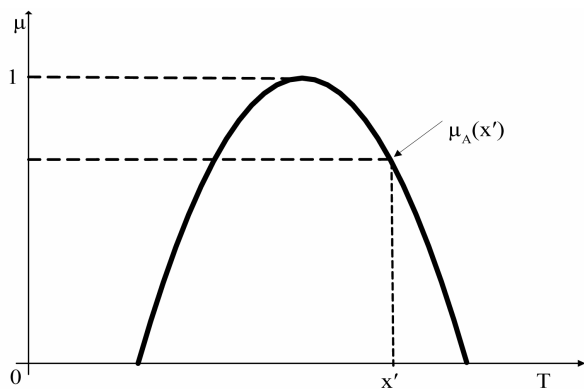


Figure 1. Type-1 membership function.

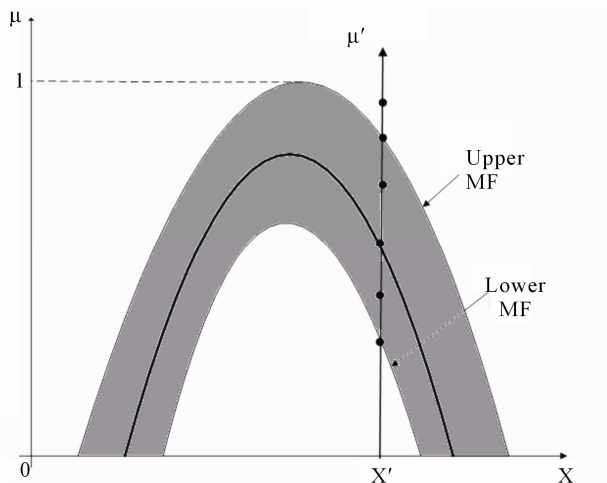


Figure 2. Blurred type-1 membership function, including discretization at $x = x'$.

those points. Doing this for all $x \in X$ we create a three-dimensional membership function: a type-2 membership function that characterizes a type-2 fuzzy set.

The basics of fuzzy logic do not change from type-1 to type-2 fuzzy sets, and in general, will not change for any type- n [21]. A higher-type number just indicates a higher “degree of fuzziness”. Since a higher type changes the nature of the membership functions, the operations that depend on the membership functions change; however, the basic principles of fuzzy logic are independent of the nature of membership functions and hence, do not change. In **Figure 3**, the general structure of a type-2 fuzzy system is presented.

Since the introduction of the concept of type-2 fuzzy set by Zadeh [20] as an extension of the concept of type-1 fuzzy set, a large number of studies have been carried out to boost the concept’s applicability, see [9-11,22,23] for details. Despite the fact that type-2 fuzzy logic is just developing, it has featured in many published articles, often with promising results [12-14, 24-27].

An interval-valued fuzzy set is a form of type-2 sets that relaxes the requirement for precise membership functions. In this case, for each x , $\mu(x)$ is an interval in $[0, 1]$. In [28], it is argued that the use of intervals is necessary to describe an expert’s degree of belief. “Many people believe that assigning an exact number to an expert’s opinion is too restrictive and that the assignment of an interval of values is more realistic” [28]. Further details on interval type-2 fuzzy set can be found in [10,14, 22,23].

Generally, a type-2 fuzzy logic system contains five components—fuzzifier, rules, inference engine, type-reducer and defuzzifier that are inter-connected, as shown

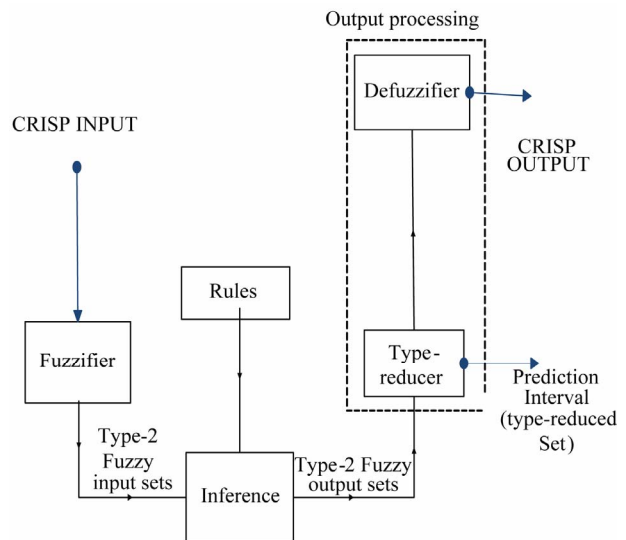


Figure 3. A type-2 fuzzy inference system.

in **Figure 3**. The general model can be viewed as a mapping from inputs to outputs and this mapping can be expressed quantitatively as $Y = f(x)$. A detailed description of each component, together with the different forms of uncertainties handled in type-2 fuzzy logic, is given in [9-10,22].

Rules are the heart of a FLS, may either be provided by experts or can be extracted from numerical data (as in ANFIS). In either case, rules can be expressed as a collection of IF-THEN statements. Consider a type-2 fuzzy logic system (T2 FLS) having p inputs

$$x_1 \in X_1, \dots, x_p \in X_p \text{ and one output } y \in Y.$$

It is assumed that there are M rules where the i^{th} rule has the form

$$R^i : \text{if } x_1 \text{ is } \tilde{F}_1^i \text{ and } \dots x_p \text{ is } \tilde{F}_p^i ,$$

$$\text{then } y \text{ is } \tilde{G}^i , i = 1, \dots, M$$

This rule represents a type-2 (T2) relation between the input space, $X_1 \times \dots \times X_p$, and the output space, Y , of the T2 FLS. Associated with the p antecedent T2 FLSs, F_k^i , are the T2-MFs $\mu_{\tilde{F}_k^i}(x_k)$, ($k = 1, \dots, p$), and associated with the consequent interval type-2 (IT2 FLS) G^i is its T2 MF $\mu_{G^i}(y)$.

3.1.1. The Proposed Type-2 Fuzzy Based Intelligence Classification Scheme

In this paper, the type-2 fuzzy logic framework is presented and utilized for identifying “question” segment in a recorded speech. The proposed type-2 FLS framework that has the ability to take care of all types of uncertainty and imprecision is presented in this section. The goal is to completely specify the FLSs using the training data, which is a unique characteristic of adaptive fuzzy systems.

Initializing a FLS involves the initialization of its antecedents, consequents and the fuzzy rules. These components of a fuzzy logic system can be initialized either from the numerical dataset or from the expert opinion. In this work, we initialized FLS from the numerical dataset

consisting of the prosodic features extracted from the recorded speech signal. In this model, we have the antecedents and consequents:

- Internal attributes are the antecedents, which, in this case, are the input parameters to the system (prosodic features)
- External attribute is the consequent which is the target class value to be predicted in this case.

A further detailed explanation of the various ways of initializing and training type-2 FLS can be found in [9]. The proposed type-2 FLS classification scheme is further shown in **Figure 4** below.

3.1.2. Training the Model with Type-2 Learning Procedures

After initializing the FLS, part of the available dataset will be used as training data. It will contain the input-output pair where the inputs are independent variables and the output is the target attribute. The objective of the training algorithm is to minimize the error function for E training Epochs.

$$E(t) = \left(\frac{f(x^{(t)}) - y^{(t)}}{y^{(t)}} \right)^2, t = 1, \dots, N$$

The steps are as follows:

- 1) Initialize all the parameters.
- 2) Set the counter of the training epoch, starting from zero.
- 3) Apply the means of the internal attribute measurements with their corresponding standard deviation to the type-1 non-singleton type-2 FLS.
- 4) Use the modified height defuzzification in a defuzzifier.
- 5) Tune the uncertain means of the antecedent membership functions and the consequents by using the Steepest Descent or Heuristic based algorithm for the error function.
- 6) Calculate error function. If $e = \text{Threshold}$ then Stop.

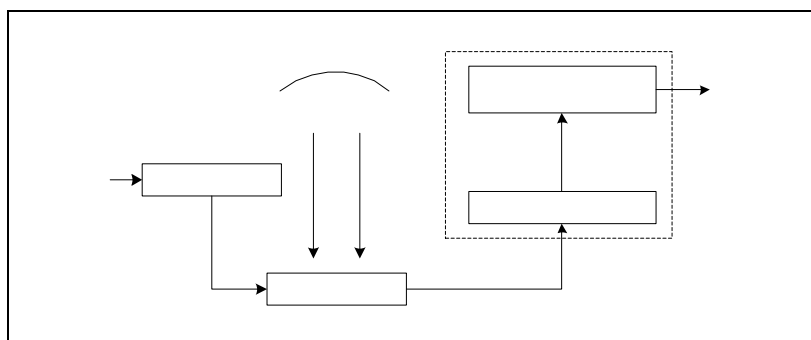


Figure 4. Proposed FLS classification framework.

Otherwise start a new epoch

For further algorithmic details on type-2 fuzzy logic systems, interested readers can consult [9-11,14,26,27, 29-32].

3.2. The Proposed Intelligent Framework Based on Sensitivity-Based Linear Learning Method (SBLLM)

In [15], the authors proposed a new learning scheme in order to both speed up and avoid local minima convergence of the existing backpropagation learning technique. This new learning strategy is called the Sensitivity-Based Linear Learning Method (SBLLM) scheme. It is a learning technique for two-layer feedforward neural networks (shown in **Figure 5**) based on sensitivity analysis, which uses a linear training algorithm for each of the two layers. First, random values are assigned to the outputs of the first layer. These initial values are then updated later based on sensitivity formulas, which use the weights in each of the layers; the process is then repeated until convergence is achieved. Since these weights are learnt by solving a linear system of equations, there is an important saving in computational time. The method also gives the local sensitivities of the least square errors with respect to input and output data, at no extra computational cost, because the necessary information becomes available without any extra calculations. This new scheme can also be used to provide an initial set of weights, which significantly improves the behavior of other learning algorithms. The full theoretical basis for SBLLM and its performance has been demonstrated in [15], which contains its application to several learning problems examples in which it has been favorably compared with several learning algorithms and well known data sets. The results have shown a learning speed that is generally faster than that of other existing methods. In addition, it can be used as an initialization tool for other well known methods, leading to significant improvements.

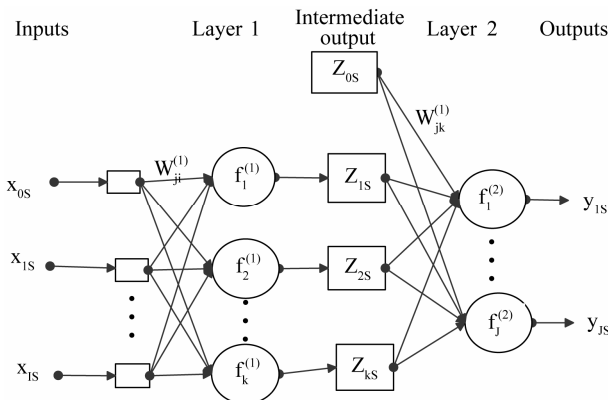


Figure 5. Two-layer feedforward neural network [15,35].

Sensitivity analysis is a very useful technique for deriving how and how much the solution to a given problem depends on data, as shown in [16,33,34] and the references therein. However, in [15] it was shown that sensitivity formulas can also be used as a novel supervised learning algorithm for two-layer feedforward neural networks that present a high convergence speed. Generally, SBLLM process is based on the use of the sensitivities of each layer’s parameters with respect to its inputs and outputs and also on the use of independent systems of linear equations for each layer to obtain the optimal values of its parameters. In addition, it gives the sensitivities of the sum of squared errors with respect to the input and output data.

The Learning Process for the Proposed SBLLM Framework

The training algorithm of the SBLLM technique can be summarized in the following algorithmic steps. More details on the workings of the SBLLM are provided in [15].

Input—The inputs to the system, which are the prosody features that make up the data (training) set (input x_{is} , and desired data, y_{js}), two threshold errors (ε and ε') to control both convergence and a step size ρ .

Output—The output results of the SBLLM system are the weights of the two layers and the sensitivities of the sum of squared errors with respect to the input and output data.

Step 0: Initialization. Assign to the outputs of the intermediate layer the output associated with some random weights $w^{(1)}(0)$ plus a small random error, that is,

$$z_{ks} = f_k^{(1)} \left(\sum_{i=0}^I w_{ki}^{(1)}(0) x_{is} \right) + \varepsilon_{ks}; \varepsilon_{ks} \rightarrow U(-\eta, \eta);$$

$$k = 1, \dots, K,$$

where η is a small number, and initialize the sum of squared errors $(Q)_{\text{previous}}$ and mean-squared errors $(MSE)_{\text{previous}}$ to some large number, where MSE measures the error between the obtained and the desired output.

Step 1: Sub-problem solution. Learn the weights of layers 1 and 2 and the associated sensitivities solving the corresponding systems of equations, that is,

$$\sum_{i=0}^I A_{pi}^{(1)} w_{ki}^{(1)} = b_{pk}^{(1)}, \sum_{k=0}^K A_{qk}^{(2)} w_{jk}^{(2)} = b_{qj}^{(2)}$$

where

$$A_{pi}^{(1)} = \sum_{s=1}^S x_{is} x_{ps}; b_{pk}^{(1)} = \sum_{s=1}^S f_k^{(1)-1}(z_{ps}) x_{ps};$$

$$p = 0, 1, \dots, I; k = 1, 2, \dots, K,$$

and

$$A_{qk}^{(2)} = \sum_{s=1}^S z_{ks} z_{qs}; b_{qj}^{(2)} = \sum_{s=1}^S f_k^{(2)^{-1}}(y_{js}) z_{qs};$$

$$q = 0, 1, \dots, K; \forall j.$$

Step 2: Evaluate the sum of squared errors. Evaluate Q using

$$Q(z) = Q^{(1)}(z) + Q^{(2)}(z)$$

$$= \sum_{s=1}^S \left[\sum_{k=1}^K \left(\sum_{i=0}^I w_{ki}^{(1)} x_{is} - f_k^{(1)^{-1}}(z_{ks}) \right)^2 + \sum_{j=1}^J \left(\sum_{r=0}^K w_{jr}^{(2)} z_{rs} - f_j^{(2)^{-1}}(y_{js}) \right)^2 \right]$$

and evaluate also the MSE.

Step 3: Convergence checking. If $|Q - Q_{\text{previous}}| < \varepsilon$ or $|\text{MSE}_{\text{previous}} - \text{MSE}| < \varepsilon'$ stop and return the weights and the sensitivities. Otherwise, continue with Step 4.

Step 4: Check improvement of Q . If $Q > Q_{\text{previous}}$ reduce the value of ρ , that is, $\rho = \rho/2$ and return to the previous position, that is, restore the weights, $z = z_{\text{previous}}$, $Q = Q_{\text{previous}}$ and go to Step 5. Otherwise, store the values of Q and z , that is, $Q_{\text{previous}} = Q$, $\text{MSE}_{\text{previous}} = \text{MSE}$ and $z_{\text{previous}} = z$ and obtain the sensitivities using:

$$\frac{\partial Q}{\partial z_{ks}} = - \frac{2 \left(\sum_{i=0}^I w_{ki}^{(1)} x_{is} - f_k^{(1)^{-1}}(z_{ks}) \right)}{f_k^{(1)}(z_{ks})}$$

$$+ 2 \sum_{j=1}^J \left(\sum_{r=0}^K w_{jr}^{(2)} z_{rs} - f_j^{(2)^{-1}}(y_{js}) \right) w_{jk}^{(2)},$$

$$k = 1, \dots, K.$$

Step 5: Update intermediate outputs. Using the Taylor series approximation in equation:

$$Q(z + \Delta z) = Q(z) + \sum_{k=1}^K \sum_{s=1}^S \frac{\partial Q(z)}{\partial z_{ks}} \Delta z_{ks} \approx 0,$$

And then update the intermediate outputs as

$$z = z - \rho \frac{Q(z)}{\|\nabla Q\|^2} \nabla Q,$$

and go to Step 1.

4. Empirical Study, Discussion and Comparative Studies

In order to carry out an empirical study, the available data sets were made use of after passing them through a preprocessing stage in order to clean the data of abnormalities or outliers (if any). To evaluate the performance of each of the modeling scheme proposed, the entire database is divided using the stratified sampling approach. 70% of the data were used for training and building each

of the models (internal validation) while the remaining 30% were reserved for testing/validation (external validation or cross-validation criterion). Generally, once the proposed models of type-2 fuzzy logic systems and the sensitivity-based linear learning method (SBLLM) have been both trained, the calibration model becomes ready for testing and evaluation using the cross-validation criterion. In order to test and evaluate the two proposed frameworks, and effectively compare their individual performances with that of the support vector machines (SVM) used in our previous study, the classification accuracy was calculated using Percent Correct measure, which is a measure of the percentage of correctly classified target classes.

4.1. Implementation Details and Parameters Settings

As for the implementation of the three methods, we made use of very few ready-made software functions. The entire coding has been largely developed by us using MATLAB, though some built-in MATLAB functions, used especially for ANN, and few others functions made available online, particularly those of SVM, have been called and used in some cases. Also part of the type-2 fuzzy logic functions made available in [9] were also used in our work. As such, the development of the new code which has represented a major effort in our research work, represents by itself a notable contributions of our paper.

In the case of the type-2 FLS-based model, the implementation process proceeded by supplying the system with the available input data sets, one sample at a time, and the rules and membership functions were automatically learned from the available input data. Gaussian membership function was used based on two different learning criteria that include least squares and back-propagation schemes. The same combination was utilized in training FLS membership function parameters. Further details on initializing, training and validating type-2 FLS have been presented in Section 3.0, and additional details could be found in [9,36,37]. The outputs from the type-2 based model are reduced to type-2 fuzzy sets, which are finally defuzzified to produce the final crisp output representing the predicted class. In this work, a type reducer algorithm, known as the Centre-of-sets (COS) algorithm, proposed by Mendel and Karnik [9,23] known as Centre-of-sets (COS) is used here as it is the most utilized in the literature [9,31,38].

As for implementation of the sensitivity-based linear learning method (SBLLM), the number of hidden neurons was set to be 100 while the activation function used was chosen to be a sigmoid (sig) activation function with the learning epoch set to 1000 or 0.001 goal error, and a

learning rate of 0.01. Regarding the support vector machines (SVM) implementation, an optimization was carried out to arrive at the best parameters that include the regularization factor $C = 450$, $\epsilon = 0.2$ while the kernel option was chosen to be Gaussian. These were all arrived at after an extensive simulation-based optimization.

4.2. Discussion and Comparative Studies

The results of performance comparisons for both training and testing sets (internal and external validation check respectively) are summarized in **Tables 2** and **3** and also pictorially represented in **Figures 6** to **9** in order to easily make at-a-glance analysis of the results. It could be easily observed from these results that, with 95 prosodic features used, whereas the SBLLM performed slightly better than the SVM, the type-2 FLS outperformed both the SBLLM and SVM classifiers which achieved a maximum accuracy of 75.61% and 75.21% of correctly classified target classes, respectively. These improved accuracy levels represent a 3.78% improvement over SVM for the type-2 FLS and a 0.5% improvement over SVM for the SBLLM.

Table 2. Training and testing results for the case with 95 features.

Classifier	Training accuracy (%)	Testing accuracy (%)
Type-2 FLS	100	78.0488
SBLLM	78.49	75.61
SVM	81.6	75.21

Table 3. Training and Testing results for the case with 65 features.

Classifier	Training accuracy (%)	Testing accuracy (%)
Type-2 FLS	100	79.5122
SBLLM	77.89	76.59
SVM	79.23	74.01

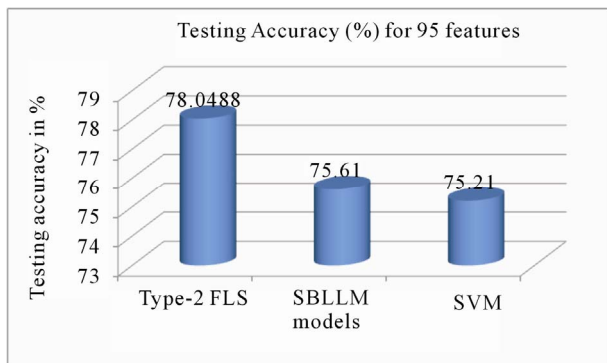


Figure 6. Pictorial representation of the testing results for the case with 95 features.

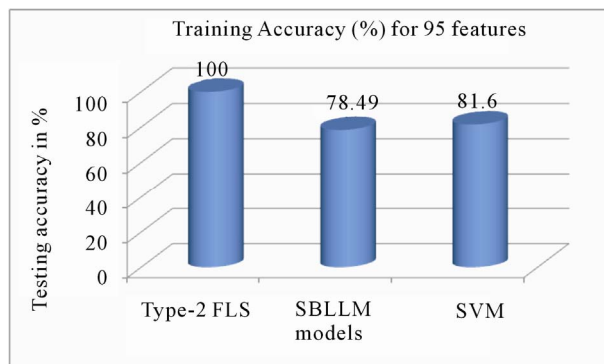


Figure 7. Pictorial representation of the training results for the case with 95 features.

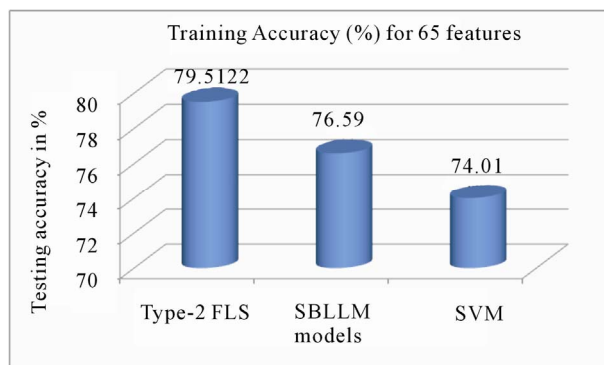


Figure 8. Pictorial representation of the testing results for the case with 65 features.

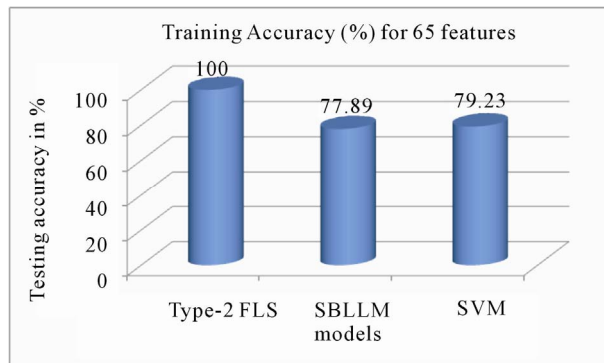


Figure 9. Pictorial representation of the training results for the case with 65 features.

However, when the total number of prosodic features used was reduced to 65, thereby increasing the amount of uncertainty, then, as **Table 3** shows, the SVM accuracy had been adversely affected whereas both type-2 FLS and SBLLM performed brilliantly well, far above the maximum of 74.01% attained by the SVM in our previous work [39], with type-2 FLS reaching 79.51% testing accuracy and SBLLM having 76.59 testing accuracy. These improved accuracy levels represent a 7.43% improvement over SVM for type-2 FLS and a 3.49% im-

provement over SVM for SBLLM.

From the above experimental results, it is clear that as the number of prosodic features decreases, the SVM suffers a loss of classification accuracy whereas the two proposed classifiers gain in accuracy. By its very nature, the type-2 FLS enjoys a certain resilience or robustness to uncertainty, whereas the SBLLM, albeit in a lesser degree than the type-2 FLS, seem to derive its robustness to uncertainty from its reliance on its sensitivities to both input and output.

Overall, it can be seen that Type-2 FLS has comparatively performed outstandingly better in all experiments carried out than the other proposed classifier (SBLLM) and the SVM used in our previous study [17,39]. Thus, these encouraging results emanating from this preliminary research based on a only a small database of questions and non-questions segments, show that the two proposed classifiers of type-2 FLS and SBLLM are potential candidates for this interesting field of research.

Furthermore, in order to confirm the significance of our results, we carried out the two-tailed t-test of significance of our results using a 0.95 level of confidence. These results are presented in **Table 4** below. The results clearly show that the improvement provided by our proposed methods over the SVM is indeed significant and not merely due to randomness.

5. Conclusion and Recommendation

In this paper, two novel approaches, namely type-2 fuzzy logic systems (type-2 FLS), and sensitivity-based linear learning machines (SBLLM), have been proposed and investigated in the detection of question/non-question segments in Arabic audio monologues using prosodic features. The improvements they brought to the classification accuracy, over previously-used techniques such as the SVM, in the face of uncertainty due mainly to the smallness of the database of monologues used, are certainly encouraging and make these two approaches strong potential candidates in similar classification tasks. From this preliminary study, the following conclusions and recommendations could be drawn:

Two modeling schemes based on type-2 FLS, and SBLLM have been investigated, developed and implemented, as predictive solutions for identifying questions/non-questions in Arabic audio monologues using pro-

sodic features. In-depth comparative studies have been carried out among these frameworks in comparison to the SVM previously-used in [17,39]. Empirical results from simulations show that type-2 FLS and SBLLM-based models outperformed SVM, with type-2 FLS taking the lead in all reported results.

With prosody-based research in the Arabic language still in its nascent state, when compared to its English language-based counterpart, it is suggested that further work be done in this area to improve upon the results achieved so far. An interesting further research direction is to explore ways in which the outputs from the different classifiers here studied can be combined into a single classifier for the question/non-questions identification task, in a framework known as an “ensemble framework” (or to technically link different classifiers together to form hybrid frameworks). It is anticipated that such combined classifier (ensemble/hybrid) would provide a powerful questions/non-questions detection tool as it would benefit from the individual advantages of its constituent classifiers while producing unique benefits of hybrid/ensemble models. Moreover, the two proposed approaches; either used individually or in an ensemble/hybrid framework, could also be extended to the classification of more than two classes. A further extension would entail the use of dialogs instead of monologues and the use of a much larger corpus for studying different speech acts in Arabic Language.

6. Acknowledgements

The authors would like to acknowledge the support provided by the Deanship of Scientific Research at King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia, under the Research Grant FT070008.

REFERENCES

- [1] I. H. Witten and E. Frank, “Data Mining, Practical Machine Learning Tools and Techniques,” 2nd Edition, Elsevier, San Francisco, 2005.
- [2] A. Selamat, S. O. Olatunji and A. A. Abdul Raheem, “A Hybrid Model through the Fusion of Type-2 Fuzzy Logic Systems and Sensitivity-Based Linear Learning Method for Modeling PVT Properties of Crude Oil Systems,” *Advances in Fuzzy Systems*, Vol. 2012, 2012, Article ID: 359429. [doi:10.1155/2012/359429](https://doi.org/10.1155/2012/359429)
- [3] www.islamway.com
- [4] E. Shriberg, *et al.*, “Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?” *Language and Speech Special Issue on Prosody and Conversation*, Vol. 41, 1998, pp. 443-492.
- [5] R. Fernandez and R. W. Picard, “Dialog Act Classification from Prosodic Features Using Support Vector Machines,” *Proceedings of Speech Prosody*, Aix-en-Pro-

Table 4. Results of statistical significance interval test.

Approach	Average recognition rate	Statistical significance interval
SVM	68.2927%	+/-6.4237%
SBLLM	75.6098%	+/-5.8874%
Type-2 FLS	76.0976%	+/-5.8874%

- vence, 11-13 April 2002.
- [6] M. Swerts, "Prosodic Features at Discourse Boundaries of Different Strengths," *Journal of the Acoustical Society of America Phonetics*, Vol. 101, No. 1, 1997, pp. 514-521.
- [7] S. Pfeiffer, "Pause Concepts for Audio Segmentation at Different Semantic Levels," *Proceedings of the ninth ACM International Conference on Multimedia*, Ottawa, 20 September 2001-5 October 2001, pp. 187-193.
- [8] C. M. Lee and S. S. Narayanan, "Toward Detecting Emotions in Spoken Dialogs," *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 2, 2005, pp. 293-303.
- [9] J. M. Mendel, "Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions," Prentice Hall, Upper Saddle River, 2001.
- [10] J. M. Mendel and R. I. B. John, "Type-2 Fuzzy Sets Made Simple," *IEEE Transactions on Fuzzy Systems*, Vol. 10, No. 2, 2002, pp. 117-127. [doi:10.1109/91.995115](https://doi.org/10.1109/91.995115)
- [11] J. M. Mendel, "Fuzzy Sets for Words: A New Beginning," *The 12th IEEE International Conference on Fuzzy Systems*, Los Angeles, 25-28 May 2003, pp. 37-42.
- [12] Q. Liang, N. N. Karnik and J. M. Mendel, "Connection Admission Control in ATM Networks Using Survey-Based Type-2 Fuzzy Logic Systems: Applications and Reviews," *IEEE Transactions on Systems, Man and Cybernetics: Part C*, Vol. 30, No. 3, 2000, pp. 329-339. [doi:10.1109/5326.885114](https://doi.org/10.1109/5326.885114)
- [13] M. H. F. Zarandi, *et al.*, "A Type-2 Fuzzy Rule-Based Expert System Model for Stock Price Analysis," *Expert Systems with Applications*, Vol. 36, No. 1, 2009, pp. 139-154.
- [14] W. Dongrui and J. M. Mendel, "A Vector Similarity Measure for Linguistic Approximation: Interval Type-2 and Type-1 Fuzzy Sets," *Information Sciences*, Vol. 178, No. 2, 2008, pp. 381-402. [doi:10.1016/j.ins.2007.04.014](https://doi.org/10.1016/j.ins.2007.04.014)
- [15] E. Castillo, *et al.*, "A Very Fast Learning Method for Neural Networks Based on Sensitivity Analysis," *Journal of Machine Learning Research*, Vol. 7, 2006, pp. 1159-1182.
- [16] A. S. Castillo, *et al.*, "A General Method for Local Sensitivity Analysis with Application to Regression Models and Other Optimization Problems," *Technometrics*, Vol. 46, No. 4, 2004, pp. 430-445. [doi:10.1198/004017004000000509](https://doi.org/10.1198/004017004000000509)
- [17] O. Khan, "Detection of Questions in Arabic Audio Monologues Using Prosodic Features," *The 9th International Symposium on Multimedia*, Taichung, 10-12 December 2007, pp. 29-36.
- [18] E. Cox, "Adaptive Fuzzy Systems," *IEEE Spectrum*, Vol. 30, No. 2, 1993, pp. 27-31. [doi:10.1109/6.208359](https://doi.org/10.1109/6.208359)
- [19] D. Dubois and H. Prade, "Fuzzy Sets and Systems: Theory and Applications," Academic Press, New York, 1982
- [20] L. A. Zadeh, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning—I," *Information Sciences*, Vol. 8, No. 3, 1975, pp. 199-249. [doi:10.1016/0020-0255\(75\)90036-5](https://doi.org/10.1016/0020-0255(75)90036-5)
- [21] S. S. Lee and K. H. Lee, "A Ranking Method for Type-2 Fuzzy Values," *Journal of Korea Fuzzy and Intelligent Systems Society*, Vol. 12, No. 4, 2002, pp. 341-346. [doi:10.5391/JKIFS.2002.12.4.341](https://doi.org/10.5391/JKIFS.2002.12.4.341)
- [22] Q. Liang and J. M. Mendel, "Interval Type-2 Fuzzy Logic Systems: Theory and Design," *IEEE Transactions on Fuzzy Systems*, Vol. 8, No. 5, 2000, pp. 535-550.
- [23] N. N. Karnik, J. M. Mendel and Q. Liang, "Type-2 Fuzzy Logic Systems," *IEEE Transactions on Fuzzy Systems*, Vol. 7, No. 6, 1999, pp. 643-658. [doi:10.1109/91.811231](https://doi.org/10.1109/91.811231)
- [24] Q. Liang and J. M. Mendel, "Equalization of Non-linear Time-Varying Channels Using Type-2 Fuzzy Adaptive Filters," *IEEE Transactions on Fuzzy Systems*, Vol. 8, No. 5, 2000, pp. 551-563. [doi:10.1109/91.873578](https://doi.org/10.1109/91.873578)
- [25] Q. Liang, and J. M. Mendel, "Overcoming Time-Varying Co-channel Interference Using Type-2 Fuzzy Adaptive Filters," *IEEE Transactions on Circuits and Systems*, Vol. 47, No. 12, 2000, pp. 1419-1428. [doi:10.1109/82.899635](https://doi.org/10.1109/82.899635)
- [26] S. O. Olatunji, A. Selamat and A. Abdurraheem, "Modeling the Permeability of Carbonate Reservoir Using Type-2 Fuzzy Logic Systems," *Computers in Industry*, Vol. 62, No. 2, 2011, pp. 147-163. [doi:10.1016/j.compind.2010.10.008](https://doi.org/10.1016/j.compind.2010.10.008)
- [27] S. O. Olatunji, A. Selamat and A. A. A. Raheem, "Predicting Correlations Properties of Crude Oil Systems Using Type-2 Fuzzy Logic Systems," *Expert Systems with Applications*, Vol. 38, No. 9, 2011, pp. 10911-10922. [doi:10.1016/j.eswa.2011.02.132](https://doi.org/10.1016/j.eswa.2011.02.132)
- [28] H. Nguyen, V. Kreinovich and Q. Zuo, "Interval Valued Degrees of Belief: Applications of Interval Computations to Expert Systems and Intelligent Control," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 5, No. 3, 1997, pp. 317-358. [doi:10.1142/S0218488597000257](https://doi.org/10.1142/S0218488597000257)
- [29] F. Liu, "An Efficient Centroid Type-Reduction Strategy for General Type-2 Fuzzy Logic System," *Information Sciences*, Vol. 178, No. 9, 2008, pp. 2224-2236. [doi:10.1016/j.ins.2007.11.014](https://doi.org/10.1016/j.ins.2007.11.014)
- [30] J. M. Mendel, "Computing with Words and Its Relationships with Fuzzistics," *Information Sciences*, Vol. 177, No. 4, 2007, pp. 988-1006. [doi:10.1016/j.ins.2006.06.008](https://doi.org/10.1016/j.ins.2006.06.008)
- [31] J. M. Mendel and H. Wu, "New Results about the Centroid of an Interval Type-2 Fuzzy Set, Including the Centroid of a Fuzzy Granule," *Information Sciences*, Vol. 177, No. 2, 2007, pp. 360-377. [doi:10.1016/j.ins.2006.03.003](https://doi.org/10.1016/j.ins.2006.03.003)
- [32] D. Wu and J. M. Mendel, "Uncertainty Measures for Interval Type-2 Fuzzy Sets," *Information Sciences*, Vol. 177, No. 23, 2007, pp. 5378-5393. [doi:10.1016/j.ins.2007.07.012](https://doi.org/10.1016/j.ins.2007.07.012)
- [33] A. C. Castillo, J. M. Guti'erez and R. E. Pruneda, "Working with Differential, Functional and Difference Equations Using Functional Networks," *Applied Mathematical Modelling*, Vol. 23, No. 2, 1999, pp. 89-107. [doi:10.1016/S0307-904X\(98\)10074-4](https://doi.org/10.1016/S0307-904X(98)10074-4)
- [34] A. S. Castillo, J. M. Guti'erez and A. Hadi, "Sensitivity Analysis in Discrete Bayesian Networks," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 26, no. 7, 1997, pp. 412-423. [doi:10.1109/3468.594909](https://doi.org/10.1109/3468.594909)
- [35] S. O. Olatunji, *et al.*, "Modeling the Correlations of

Crude Oil Properties Based on Sensitivity Based Linear Learning Method,” *Engineering Applications of Artificial Intelligence*, Vol. 24, No. 4, 2011, pp. 686-696. doi:10.1016/j.engappai.2010.10.007

- [36] J. M. Mendel and R. I. B. John, “Type-2 Fuzzy Sets Made Simple,” *IEEE Transactions on Fuzzy Systems*, Vol. 10, No. 2, 2002, pp. 117-127. doi:10.1109/91.995115
- [37] O. Castillo, “Type-2 Fuzzy Logic in Intelligent Control Applications, Studies in Fuzziness and Soft Computing Series, Vol. 272, Springer, Berlin, 2012.
- [38] N. N. Karnik and J. M. Mendel, “Centroid of a Type-2 Fuzzy Set,” *Information Sciences*, Vol. 132, No. 1-4, 2001, pp. 195-220. doi:10.1016/S0020-0255(01)00069-X
- [39] K. Omair, G. A.-K. Wasfi and C. Lahouari, “A Preliminary Study of Prosody-Based Detection of Questions in Arabic Speech Monologues,” *The Arabian Journal for Science and Engineering*, Vol. 35, No. 2, 2010, pp. 167-181.