

Using bacterial foraging optimization approach to optimize the learning of hidden...

[Science](#), [Biology](#)



Hidden Markov model using the bacterial foraging optimization algorithm for speech recognition scientists in the area.

The swarm of microorganisms “ Bacteria” S acts as takes after:

Bacteria are discretionarily spread in the map of nutrients.

Bacteria travel towards high-nutrient regions on the map. Those get their nourishment inadequate, expanded long and will split into two equal parts at a reasonable temperature. Besides those situated in a low nutrient district will scatter and the individuals who located in a toxic area will die.

Bacteria located in the promising districts of their environment will attempt to pull in others microorganisms by producing chemical attractants.

Bacteria are situated in the most elevated nutrient districts.

Bacteria scatter as to search for new nutrient districts in their environment.

Bacteria foraging behavior searching conduct incorporates four primary advances: Chemo-taxis (tumble and swimming), swarming, Reproduction and Elimination-dispersal.

The Proposed BFOA/HMM Training

In this study, we explore the application of BFOA in order to optimize the learning of the HMMs. As other swarm intelligence algorithms, BFOA is inspired from the social and cooperative actions found in nature. In fact, the optimization process here consists in the way bacteria rummage around for the high nutrient regions. The application of BFOA is various in real-world

optimization problems and therefore gained the attention of the researchers in the field.

Suppose that we want to find the minimum of $J(\mathbf{q})$ where $\mathbf{q} \in \mathbb{R}^p$ (i. e. \mathbf{q} is a p -dimensional vector of real numbers), and no measurements or an analytical description of the gradient $\nabla J(\mathbf{q})$ is available.

The definition of a chemotaxis step to be a tumble followed by a tumble or a tumble followed by a run is as follows:

j be the index for the chemotaxis step, k the index for the reproduction step, and l the index of the elimination-dispersal event. The parameters are described in the Table below:

Table 1: Description of Parameters

- Parameter Name Description
- p The dimension of the search space
- S Total number of bacteria in the population
- N_c The number of chemo-tactic steps
- N_s The swimming length
- N_{re} The number of reproduction steps
- N_{ed} The number of elimination-dispersal events
- P_{ed} Elimination-dispersal probability,
- $C(i)$ The size of the step taken in the random direction specified by the tumble
- $J(i, j, k, l)$ denote the cost at the location of the i th bacterium

-

i (j, k, l) the indices are dropped and refer to the i th bacterium position as θ_i . Note that we will interchangeably refer to J as being a “cost” (using terminology from optimization theory) and as being a nutrient surface (in reference to the biological connections). A brief description of the four prime steps in BFOA is presented below:

Chemo-taxis

Biologically the movement of E. Coli bacterium is restricted into two different ways, either it can swim for a period of time in the same direction or it may tumble for the entire lifetime. In the BFOA, a step in a random way speaks to a tumble and a step with similar course indicates a run, which simulates the chemo-taxis procedure is used in Eq. (6). θ_i

j presents the position of the i th bacterium in the j th chemo-taxis step, $C(i)$ is the step length during the j th chemotaxis and $\varnothing(i)$ is a unit vector which stands for the swimming direction after a tumble. It can be generated by Eq. (7), where Δ_i is a randomly produced vector with the same dimension of the problem:

$$\theta_i$$

$$(j+1) = \theta_i$$

$$j + C(i) \cdot \varnothing(i) \quad (6)$$

$$\varnothing(i) = (\Delta(i)) / \sqrt{\Delta}$$

$T(i)\Delta(i)$ (7)

In each chemotaxis step, the bacterium generated a tumble direction firstly. Then the bacterium moves in the direction using Eq. (6). If the nutrient concentration in the new position is higher than the last position, it will run one more step in the same direction. This procedure continues until the nutrient get worse or the maximum run step is reached. The maximum run step is controlled by a parameter called N_s .

Swarming

Fascinating group behavior has been watched for a few motile species of bacteria including *E. coli* and *S. Typhimurium*, where mind-boggling and stable spatiotemporal patterns (groups) are formed in the nutrient medium. A group of *E. Coli* cells orchestrate themselves in a voyaging group by moving up the supplement nutrient when put in the midst of a semisolid environment with a single nutrient chemo-effector. The cells when stimulated by a succinate, discharge an attractant aspartate, which encourages them to total into gatherings and thus move as concentric patterns of swarms with high bacterial density.

BFOA recreates this social conduct by representing the combined cell-to-cell attraction and repelling effect that can be modeled as:

$$J_{cc}(\theta, P(j, k, l)) = \sum_{i=1}^n$$

s_{ij}

$i(\theta, \theta)$

$i(j, k, l) \quad (9)$

$J_{cc}(\theta, P(j, k, l)) = \sum_{i=1}^S$

$s \left[-d_{att} \exp(-w_{att} \sum_{m=1}^p$

$p \left(\theta_m - \theta_m$

$i)$

$2)] + \sum_{i=1}^S$

$s \left[-d_{rep} \exp(-w_{rep} \exp \sum_{m=1}^p$

$p \left(\theta_m - \theta_m$

$i)$

$2)] \quad (10)$

$J_{cc}(\theta, P(j, k, l))$ is the objective function value to be added to the actual objective function (to be minimized) to present a time-varying objective function, S is the total number of bacteria, p is the number of variables to be optimized, which are present in each bacterium and $\theta = (\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(p)})$

T is a point in the p -dimensional search domain. d_{att} , w_{att} , d_{rep} , w_{rep} are different coefficients that should be chosen properly [2, 27].

Reproduction

For every N_c time of chemotaxis steps, a reproduction step is taken in the bacteria population. The bacteria are sorted in descending order by their

nutrient obtained in the previous chemotaxis processes. Bacteria in the first half of the population are regarded as having obtained sufficient nutrients so that they will reproduce. Each of them splits into two (duplicate one copy in the same location). Bacteria in the residual half of the population die, and they are removed out from the population. The population size remains the same after this procedure. For bacterial, a reproduction step takes place after all chemo-taxis steps. Now health of i th bacterium:

J_{he}

$i = \sum_{j=1}^n$

$(N_{c+1}) \times J(i, j, k, l)$ (11)

Elimination and Dispersal

In nature, the changes of environment where population lives may affect the behaviors of the population. For example, the sudden change of temperature, nutrient concentration and the flow of water. All these may cause bacteria in the population to die or move to another place. To simulate this phenomenon, eliminate-dispersal is added in the BFOA.

After every N_{re} time of reproduction steps, an eliminate-dispersal event happens. For every bacterium, an arbitrary number is generated between 0 and 1. If the random number is less than a predetermined parameter, known as P_e , the bacterium will be eliminated, and a new bacterium is generated in the environment. In BFOA, the dispersion event happens after a certain number of reproduction processes and some bacteria are chosen depending

upon the probability to be killed and moved to another position within the environment.

Generally, in the problem of HMM learning, the individuals are HMM. It is necessary to encode the HMM to present the bacterium at j th chemotaxis on which the BOFA operators will be applied. In the second case, the BFOA is used to calculate the w , μ and Σ parameters.

The population of size S was created randomly. The most natural encoding is to build the bacteria by reorganizing all the coefficients of the HMM. The simplest way is to juxtapose all rows of all matrices. Thus, a coding in real numbers is obtained while respecting the constraints related to the HMM.

Recognition

A discriminated model is used to do Recognition, where learning will be associated with each word learned in HMM. Recognition will be done by calculating for each known HMM its probability of generating the word to recognize. The word will be recognized that the associated HMM obtained a maximum score.

The HMM who has the highest probability of generating the input data is selected in decision stage. In our experiment, the decision is performed using the Viterbi algorithm.

The Viterbi algorithm is used to find the best state sequence q_1

*, q_2

*, ..., q_T

*given the observation sequence $O = O_1, O_2, \dots, O_T$. The most elevated probability along a solitary path, which represents the primary k observations and closures in S_i at time k , is characterized as:

$$\delta_k(i) = \max_{\tau} (q_1, q_2, \dots, q_{(k-1)}) \cdot P[q_1, q_2, \dots, q_k = S_i, O_1, O_2, \dots, O_k | \lambda] \quad (12)$$

It can be induced that:

$$\delta_{(k+1)}(j) = \max_i [\sum_i a_{ij} \delta_k(i) + b_j(O_{(k+1)})] \quad (13)$$

The observation probability distribution $b_j(O_k)$ is a Gaussian mixture likelihood function as mentioned above.

EXPERIMENTAL RESULTS AND DISCUSSION

The performance analysis of various speech recognition systems is evaluated. Speech recognition has a big major in turning into a critical factor of association amongst human and computer. A successful speech recognition system has to determine features not only present in the input pattern at one point in time but also features of input pattern changing over time. The common performance measure is word blunder rate. This measure is processed by comparing a reference transcription output by the speech recognizer. From this comparison it is possible to compute the number of errors, which typically belongs to three categories:

Insertions I (when in the output of ASR it is present a word not present in reference), deletions D (a word is missed in ASR output) and substitutions S (a word is confused with another word).

Database

In order to test the proposed system, a multi-speakers database was used for a speech recognition task. Therefore, we used the Clemson University Audio-Visual Experiments (CUAVE) database [29]. It was created by the Digital Speech and Audio Processing Group at Clemson University. For research purposes, they distribute it on DVD royalty-free, which is an audio-visual database that contains over 7, 000 utterances of both connected and isolated digits recorded by 36 individual speakers (17 female and 19 male) and 20 pairs of speakers. It is aimed at testing multi-speaker solutions. It includes also both still and moving speakers in order to be robust to speaker movement.

The database contains around 3 hours of speech recorded through a Mini DV camera. The video was then compressed into MPEG-2 files (stereo audio at a 44 kHz sampling rate, 16-bit). In addition, it includes audio files checked for synchronization (mono rate of 16 kHz, 16-bit) and annotation files.

Experimental results

As a standard procedure in evaluating machine learning techniques, the dataset is split into training and test sets. Therefore, we have used $\frac{2}{3}$ of the data for the learning stage and there maining $\frac{1}{3}$ to test the effectiveness of our ASR system.

The proposed ASR system that uses RASTA-PLP method as audio features, and a BFOA/HMM for the speech modeling, was implemented as described in the previous sections. The whole programming was implemented in MATLAB.

As a result, we obtain a matrix of 27 parameters by integrating the first and

second derivative of the parameters,. The BFOA/HMMs recognizers were built using the Hidden Markov Model Toolkit (HTK) [30].

Various kinds of the instance with different BFOA control parameters have been solved with our algorithm, in order to evaluate the performance of the proposed system. Each instance is run 15 times with a different number of the mixture, a different crossover probability values between 0.4-0.9, and the value 0.01 is kept for the mutation probability, also the used maximum number of iteration for EM algorithm is 40.

We can clearly notice from Table 2 that the results are varied according to the parameters training of the BFOA. However, with $S=20$, $N_c=21$ and $M=4$, the average maximum likelihood is the highest one. Therefore, these parameters value were chosen to be used in the proposed BFOA/HMM.

After several tests over the training and test data, we find that the fitness function in BFOA is better with 21 iterations, which yields to better results for optimizing HMM parameters. In figure 3, we can notice also that the BFOA fitness function increase faster, substantially at the beginning iterations.

The experimental results of the proposed ASR system over a range of noise levels using these two models. In order to simulate various noise conditions, the artificial white Gaussian noise was added.

The experiment was conducted under a mismatched condition, where the recognizers were trained at 20dB SNR, and acoustic white Gaussian noise, ranging from -5dB to 20dB in steps of 5 dB SNR, was added to the test data.

We performed evaluation and comparison of the performance of the system by calculating the average recognition rate for each SNR levels, overall utterances for each word. Moreover, the HMM model is trained by using the traditional method (Baum- Welch algorithm). The Figure 3 shows that in most of the cases, the recognition rates obtained with the proposed BFOA/HMM system provides better results in compare is onto those obtained by the HMM-based system, and the percentage increase amount from 1. 4% to 15. 3%, along with the increase in the size of the population.

CONCLUSION

Automatic Speech Recognition is viewed as an essential piece of human-computer interfaces that are conceived to utilize discourse, among different means, to accomplish regular, inescapable figuring. The state of ASR lacks vigor, to channel and environment noise continues to be a major impediment.

ASR system takes a human discourse as an information and requires a series of words as yield. The issue of consequently perceiving discourse with the assistance of a computer is a troublesome issue, and the purpose behind this is the unpredictability of the human dialect. Lack of linguistic corpora for dialect language models seems to be a difficulty in implementing an ASR. This stems from trouble in gathering sentences with vernaculars. The extent of the issue ought to be widened into larger vocabularies, continuous speech.

We presented in this paper an approach for an ASR system using BFOA/HMM with a Gaussian mixture density for modeling the Arabic speech. In order to further optimize the solution found by the Baum-Welch algorithm.

Based on the obtained results, we can conclude that the system modeled by HMM and trained by our BFOA/HMM have higher rates of recognition than the HMM trained by the Baum-Welch algorithm. Moreover, the classification results of Baum-Welch is greatly dependent on the initialization of the parameters in contrast to the BFOA, which gives stable results independently of the initial values.

As perspective, we are planning to cover more issues about improving the performance of the system, by enhancing the size of our database and increasing the number of speakers. In addition, we intend to test our system with other alternatives methods of fusion.