Question Classification for Automatic Question-Answering in Agriculture Domain

Cho Zin Oo, Ye Kyaw Thu, Hlaing Myat Nwe, Hnin Aye Thant

***Abstract*— A robust conventional Question-Answering (QA) system is necessary for agriculture in Myanmar and not available yet. Most of the QA systems utilize the algorithms of Artificial Intelligence and Machine Learning to get the required responses. A usual QA system has two components: question classification, and answer extraction. In this paper, we focus on the question classification that is needed for a robust domain specific QA system targeting agriculture domain. We proposed a two-step approach to classify questions: first, defines the tag of each word of the question by using Conditional Random Fields (CRFs), Hidden Markov Model (HMM) and Ripple Down Rules (RDR), and then the defined tags were used in the question classification and were classified using K-Nearest Neighbor (K-NN), Naive Bayes, Decision Tree, Random Forest, Support Vector Machines (SVM) and eXtreme Gradient Boosting (XGBoost). We collected 10,853 questions from available resources associated to agriculture in Myanmar. Evaluation results on tagging experiment showed that RDR approach gave the highest F1-score of 0.96 among the three tagging approaches and classification results of K-NN, Decision Tree and SVM also gave the highest F1-score of 0.88, compared to other classification algorithms by using the tags defined by RDR.**

***Index Terms*—Question Tagging, Question Classification, Agriculture Domain in Myanmar, Tagging Approaches, Machine Learning Methods**

I. Introduction

Question-answering (QA) is an active research area in natural language processing (NLP). QA system is also called as human-computer interaction [1]. The accurate answers should be responded if the user queries the machine. There is no currently QA system for agriculture in Myanmar. Myanmar is a cultivated country and about 70% of the population depends on agriculture. Agriculture is the key to develop the country. Farmers need to know informations in time for preventing the losses in the yield and quantity of the agricultural product.

In this paper, we present question classification that is needed for the QA system of agriculture domain in Myanmar as the contribution. Our proposed system includes function tagging and classification. Function tagging is a preprocessing step for classification. We contribute this function tagging by using manually defined tags such as Veg (Vegetable), Fru (Fruit), Flo (Flower), Cer (Cereal), OilS (Oil Seed), Bea (Bean), IndRC (Industrial Raw Crop), Other, nolabel, BacD (Bacterial Diseases), PlaN (Plant Nutrition), Gro (Cultivation), AgrK (Agricultural Knowledge) and Soil. In the task of function tagging, three tagging approaches are used. Machine learning approaches are used to classify the question into one of the predefined categories: AgrK, BacD, Gro, PlaN and Soil. When classifying the questions, we compared well known classification algorithms such as K-Nearest Neighbor (K-NN), Naive Bayes, Decision Tree, Random Forest, Support Vector Machines (SVM) and eXtreme Gradient Boosting (XGBoost) to select which algorithm is the most suitable with our agriculture related data.

Cho Zin Oo is with the NLP Lab., University of Technology (Yadanarpon Cyber City), Myanmar.

Ye Kyaw Thu is with National Electronics and Computer Technology Center, Thailand.

Corresponding Authors: chozinoo78@gmail.com and yktnlp@gmail.com

Manuscript received August 8, 2021; accepted October 8, 2021; revised October 20, 2021; published online October 31, 2021.

II. Related Work

There are many QA systems for English language and further languages. Different methods are used with different resources according to their levels for different systems. In the paper [2], Hasangi Kahaduva et al. developed a QA system for the travel industry. This paper presented a comprehensive QA system. It consisted of two main steps: defining the question type and searching the Knowledge Base (KB) to find the answer of the classified question. A machine learning approach was used to identify questions, and rule-based approach was used to search the knowledge base for the answer. Joseph et al. proposed a text classification system with semantic characteristics. They used the TF-IDF model in conjunction with the word2vec model. They showed the performance result of TF-IDF, word2vec in combination with TF-IDF weighted with stop words and without stop words [3]. In the paper [4] Manmita Devi et al. proposed a quality assurance system based on ontology in agriculture to answer queries in natural language. This system generated SPARQL from natural language queries, where SPARQL is the standard query language for using RDF data. This system used a combination of NLP and Semantic Web technologies to answer the questions. In the paper [5], Ji Ho Park et al. proposed the comparison of one-step and two-step classification for abusive language detection on twitter. Their research explored a two-step approach for classification on abusive language and then classified into specific types and compared it with one-step approach of doing one multi-class classification for detecting sexist and racist languages. In the paper [6], K. S. Kalaivani et al. proposed different machine learning algorithms for the classification of Health data set using POS based machine learning techniques. This paper didn't directly classify the sentiment reviews. POS (Part-of-Speech) tagging was done firstly as the preprocessing stage. And then, they classified sentiment reviews by using POS tags as the features. Three supervised machine learning algorithms such as Naive Bayes (NB), Support Vector Machines (SVM) and Maximum Entropy (ME) were used to classify reviews.

III. Approach

This section describes one-step and two-step classification approaches. Usually, one-step classification approachs are used. One-step approach means that the input question is directly classified, but two-step approach means that the input question is not directly classified. In two-step approach, the input question is tagged firstly. And then, classify these input question by using the tags defined by the first step. In this paper, we proposed two-step classifiaction approach. In the experiment, we show that our two-step approach is better than one-step.

A. *One-Step Question Classification*

In one-step question classification, the class labels we classified are BacD, PlaN, Gro, AgrK and Soil. Their brief descriptions and sample questions can be seen in Table I. The overview of one-step question classification system can be seen in Figure 1.

TABLE I: Sample Questions of One-Step Approach with Class Label for Agriculture in Myanmar

Label	**Brief Description**	**Sample Questions**
BacD	Bacterial Disease	ကျွန်တော့် ငရုပ်ပင် တွေ မှာ ပျပိုး တွေ ကျနေ တယ် ဘယ်လိုလုပ် ရ လဲ (How to control aphid on my chile plants?)
PlaN	Plant Nutrition	ကြက်သွန် အတွက် အပင်အားဆေး ဘာ ကောင်းလဲ ပြောပြပါ (Please tell me the herbal medicine to improve the growth of onion plants.)
Gro	Cultivation	လျှော်ဖြူပင် စိုက်ပျိုးနည်း သိချင်တာပါ (How to grow hemp?)
AgrK	Agriculture Knowledge	နှမ်း က ဘယ်အချိန် မှာ စိုက်ရတာလဲ (When do the sesame grow in?)
Soil	Soil	နဂါးမောက် က ဘယ်လိုမြေမျိုး မှာ စိုက်ရတာလဲ (What kind of soil does dragon fruit grow in?)

As Figure 1, the input question will be classified as the following example. If an input question

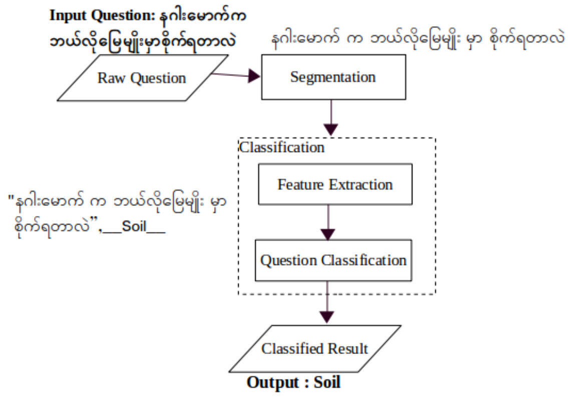

Fig. 1: The Overview of the One-Step Question Classification System

is "နဂါးမောက်ကဘယ်လိုမြေမျိုးမှာစိုက်ရတာလဲ" (What kind of soil does dragon fruit grow in?), we first segment this input question such as "နဂါးမောက် က ဘယ်လိုမြေမျိုး မှာ စိုက်ရတာလဲ". This segmented question is input for question classification. The input question format for one-step question classification is "နဂါးမောက် က ဘယ်လိုမြေမျိုး မှာ စိုက်ရတာလဲ",__Soil__ ("What kind of soil does dragon fruit grow in",__Soil__), whose class label is Soil. In this question, the user wants to know the types of soil to plant the dragon fruit. The type of this example question is Soil. Therefore the class label "Soil" will be displayed as the classification result.

B. *Two-Step Question Classification*

The overview of our two-step question classification system can be seen in Figure 2. There are two steps for identifying questions. In the first step, we define tags for the questions using CRFs, HMM, RDR tagging approaches. In the second step, we classify the tagged questions from the first step by using machine learning approaches: K-NN, Naive Bayes, Decision Tree, Random Forest, SVM and XGBoost. Tagset and question format that are used for two-step question classification are described belows.

1) *Tagset*

This section describes tagset used for function tagging, which are Veg, Fru, Flo, Cer, OilS, Bea, IndRC, Other, nolabel, BacD, PlaN, Gro, AgrK and Soil. Their brief descriptions and examples can be seen in Table II.

2) *Question Classes*

In two-step question classification, the class labels we classified are BacD, PlaN, Gro, AgrK and Soil. Their brief descriptions, sample questions and sample tagged questions can be seen in Table III. In sample tagged questions, Myanmar words "ငရုတ်ပင်" and "ကြက်သွန်" ("Chile plant and Onion" in English) are the types of vegetables. Therefore, these words are tagged as the "Veg" such as ငရုတ်ပင်/Veg and ကြက်သွန်/Veg. The all other words are tagged by using the tags and its words in Table II.

TABLE II: Tag-set of Questions for Agriculture in Myanmar

Tag	Brief Description	Examples Words
Veg	Vegetable	ငရုတ်ပင်၊ ခရမ်းချဉ်သီး၊ ကြက်သွန်၊ ... (Chile plant, Tomato, Onion, ...)
Fru	Fruit	နဂါးမောက်သီး၊ ပန်းသီး၊ လိမ္မော်သီး၊ ... (Dragon Fruit, Apple, Orange, ...)
Flo	Flower	စပယ်ပန်း၊ သစ်ခွပန်း၊ နှင်းဆီပန်း၊ ... (Jasmine, Orchid, Rose, ...)
Cer	Cereal	စပါး၊ ပြောင်း၊ ဂျုံ၊ ... (Rice, Corn, Wheat, ...)
OilS	Oil Seed	နှမ်း၊ မြေပဲ၊ နေကြာ၊ ... (Sesame, Peanut, Sunflower, ...)
Bea	Bean	မတ်ပဲ၊ ပဲတီစိမ်း၊ ကုလားပဲ၊ ... (Mung bean, Green gram, Chick pea, ...)
IndRC	Industrial Raw Crop	ကြံ၊ လျှော်ဖြူပင်၊ ဝါ၊ ... (Sugarcane, Hemp, Cotton, ...)
Other	Other	ကွမ်း၊ ကွမ်းသီး၊ ငရုတ်ကောင်း၊ ... (Betel, Betel-nut, Peppercorn, ...)
nolabel	The words that do not affect in the question classification	ကျွန်တော်၊ ကျေးဇူးပြု၍၊ ပြောပြပါ၊ သိချင်တာပါ၊ ... (I, Please, Tell me, Want to know, ...)
BacD	Bacterial Disease	ပျပိုး၊ လေးပိုး၊ ငမြောင်တောင်ပိုး၊ ... (Aphid, Thrips, Faw, ...)
PlaN	Plant Nutrition	ရွက်ဖျန်းအားဆေး၊ အပင်အားဆေး၊ အထွက်တိုးမြေပြင်မြေဩဇာ၊ ... (Foliar, Hormones, Gybsum, ...)
Gro	Cultivation	ဘယ်လိုပျိုးထောင်၊ ဘယ်လိုစိုက်၊ ... (How to cultivate, How to grow, ...)
AgrK	Agriculture Knowledge	ဘယ်ရာသီ၊ ဘယ်အချိန်၊ ကိုင်းဆက်နည်း၊ ... (Which season, When, Grafting, ...)
Soil	Soil	ဘယ်လိုမြေမျိုး၊ မြေပြုပြင်နည်း၊ ... (What kind of soil, How to prepare the land, ...)

TABLE III: Sample Questions of Two-Step Approach with Tags and Class Labels for Agriculture in Myanmar

Label	Brief Description	Sample Questions	Sample Tagged Questions
BacD	Bacterial Disease	ကျွန်တော့် ငရုတ်ပင် တွေ မှာ ပျပိုး တွေ ကျနေ တယ် ဘယ်လိုလုပ် ရ လဲ (How to control aphid on my chile plants?)	ကျွန်တော့်/nolabel ငရုတ်ပင်/Veg တွေ/nolabel မှာ/nolabel ပျပိုး/BacD တွေ/nolabel ကျနေ/nolabel တယ်/nolabel ဘယ်လိုလုပ်/nolabel ရ/nolabel လဲ/nolabel
PlaN	Plant Nutrition	ကြက်သွန် အတွက် အပင်အားဆေး ဘာ ကောင်းလဲ ပြောပြပါ (Please tell me the herbal medicine to improve the growth of onion plants.)	ကြက်သွန်/Veg အတွက်/nolabel အပင်အားဆေး/PlaN ဘာ/nolabel ကောင်းလဲ/nolabel ပြောပြပါ/nolabel
Gro	Cultivation	လျှော်ဖြူပင် စိုက်ပျိုးနည်း သိချင်တာပါ (How to grow hemp?)	လျှော်ဖြူပင်/IndRC စိုက်ပျိုးနည်း/Gro သိချင်တာပါ/nolabel
AgrK	Agriculture Knowledge	နှမ်း က ဘယ်အချိန် မှာ စိုက်ရတာလဲ (When do the sesame grow in?)	နှမ်း/OilS က/nolabel ဘယ်အချိန်/AgrK မှာ/nolabel စိုက်ရတာလဲ/nolabel
Soil	Soil	နဂါးမောက် က ဘယ်လိုမြေမျိုး မှာ စိုက်ရတာလဲ (What kind of soil does dragon fruit grow in?)	နဂါးမောက်/Fru က/nolabel ဘယ်လိုမြေမျိုး/Soil မှာ/nolabel စိုက်ရတာလဲ/nolabel

Our two-step question classification system will classify the input question as the following example. If an input question is "နဂါးမောက်ကဘယ်လိုမြေမျိုးမှာစိုက်ရတာလဲ" (What kind of soil does dragon fruit grow in?), we first segment this input question such as "နဂါးမောက် က ဘယ်လိုမြေမျိုး မှာ စိုက်ရတာလဲ". The segmented input question is tagged for question classification such as "နဂါးမောက်/Fru က/nolabel ဘယ်လိုမြေမျိုး/Soil မှာ/nolabel စိုက်ရတာလဲ/nolabel". In this example, Myanmar word "နဂါးမောက်" ("dragon fruit" in English) is the type of fruits. Therefore, this word was tagged as the "Fru". Myanmar word "က" ("stop word" in English) was tagged as the "nolabel" because this word does not affect in the question classification. The next word "ဘယ်လိုမြေမျိုး" ("What kind of soil" in English) is the type of Soil. Therefore, this word was tagged as the "Soil" and the words "မှာ" and "စိုက်ရတာလဲ" ("in and grow?" in English) were also tagged as the "nolabel" because these words do not affect in the question classification.

Then, classify the above tagged question using the only tags such as "Fru nolabel Soil nolabel nolabel",_Soil_, whose class label is Soil. This example question in Figure 2 is the same as the question that expressed in Figure 1. Therefore, this system will also display the class label "Soil" as the classification result.

IV. Function tagging

Function tagging is a process of assigning semantic categories such as Veg, Fru, Flo, Cer, OilS, Bea, IndRC, Other, nolabel, BacD, PlaN, Gro, AgrK and Soil to each word in

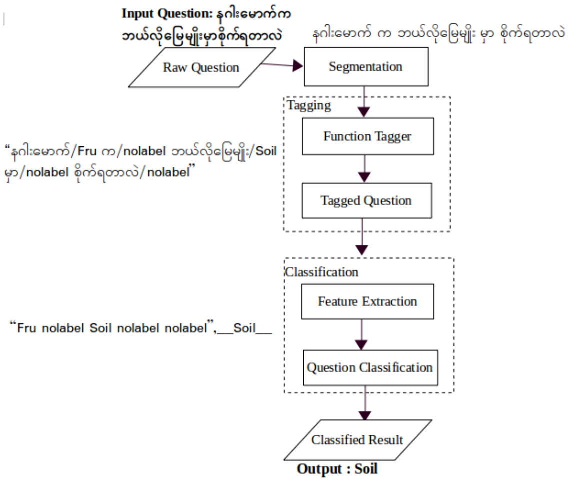

Fig. 2: The Overview of the Two-Step Question Classification System

the text questions. In this section, tagging methodologies used in the experiments are described.

A. Conditional Random Fields (CRFs)

Linear Conditional Random Fields (CRFs) [7] are models that study the dependencies between predicted segmentation labels, which are necessary for state transitions in finite state sequence models and can effectively integrate knowledge of the field in segmentation. The model estimates the following probability of a sequence of labels: Y = y_1, ..., y_T of a particular character string W = w_1, ..., w_T.

$$P_\lambda(Y/W) = \frac{1}{Z(W)} exp(\sum_{i=1}^{T} \sum_{k=1}^{\lambda} \lambda_k f_k(y_{(t-1)}, W, t)) \quad (1)$$

where $Z(W)$ is a normalization term, f_k is a feature function, and λ is a feature weight vector.

B. Hidden Markov Model (HMM)

The HMM is a probabilistic sequence model of words, letters and tags [8]. For a given sentence, the tag sequence is choosed by the HMM tagger that maximizes:

P(word/tag)*P(tag/previous tags)

A tag sequence is generally selected by the HMM tagger rather than for a word. In this approach, we compute the most potential sequence of tags T= (t_1, t_2, ..., t_n) for a sequence of words in the sentence W = (w_1, w_2, ..., w_n). A graphical representation of a HMM can be seen in Figure 3.

In the Figure 3, the words S=S_1, S_2, ..., $S_{(n-1)}$, S_n form a Markov chain, but this sequence of words is not observed (i.e.hidden). The X_1, X_2, ..., $X_{(n-1)}$, X_n are observable variables (i.e. output) of the Markov chain. Conditional dependence relations of words are indicated by the horizontal and vertical arrows.

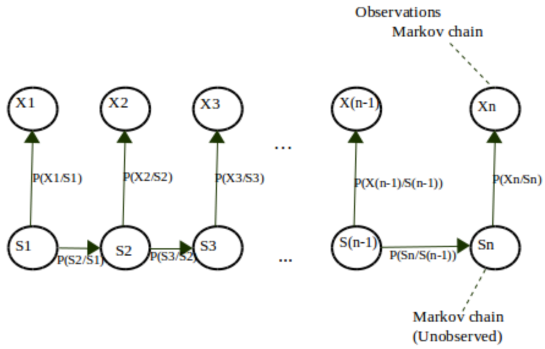

Fig. 3: Graphical Representation of a Hidden Markov Model

C. Ripple Down Rules (RDR)

Ripple-Down Rules (RDR) [9] [10] is an error-driven method that uses transformation rules to build a Single Classification Ripple Down Rules (SCRDR) tree automatically for tagging. A binary tree classification of SCRDR for tagging can be seen in Figure 4.

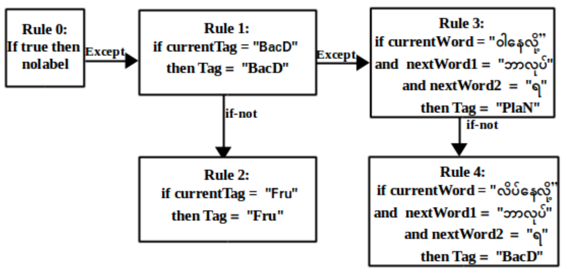

Fig. 4: A Binary Tree Classification of SCRDR

For example with the SCRDR tree in the Figure 4, given a case "ပဲတီစိမ်း/Bea တွေ/nolabel ဝါနေလို့/BacD ဘာလုပ်/nolabel ရ/nolabel မလဲ/nolabel", where tagging was done for each word according to this SCRDR tree. As this case, "ဝါနေလို့" is the current word, next word1 is "ဘာလုပ်" and next word2 is "ရ", the case satisfies the conditions of the rules at Rules (0), (1) and (3), it then passes to the other nodes. As the case satisfies the condition at Rule (3), the tag for "ဝါနေလို့" is concluded as "PlaN". As the user want to know about plant nutrition, the tag of Myanmar word "ဝါနေလို့" ("less plant nutrition" in English) was PlaN. The all other words were tagged by using the rules of the complete SCRDR tree. In this paper, we express some part of the SCRDR tree.

SCRDR is assessed by moving access to Rule (0) (default root node) of the tree as shown in Figure 4. Rule (1) is an exception of Rule (0). Rule (2) is a child of Rule (1), otherwise, an exception rule for Rule (0). Furthermore, Rules (3) and (4) are the exception rules of Rule (1).

V. Question Classification

After tagging the questions, we need to classify these questions to know their class. Multi-class and multi-label classification are performed in our classifier. Thus, the questions containing several semantic relations are handled by the classifier which is often the case in real life situations.

No matter what learning algorithm or approach is applied, text-based features remain important for the classification task. We used TF-IDF (term frequency–inverse document frequency) [11] feature extraction for classification. This model can simplify the extraction process and is easy to understand. TF-IDF is a term weighting technique that indicates the importance of a term in a document to represent textual data. It compares the frequency of words described in an individual document and throughout the document. TF-IDF bases on bag-of-words model (BoW) and does not require position in text, semantics, matches in different documents, etc.

A. K-Nearest Neighbor

K-Nearest Neighbor (K-NN) algorithm is one of the simplest machine learning algorithms [12]. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is the most similar to the available categories. It is also a lazy learning algorithm because it does not learn from the training set immediately. Instead it stores the dataset, it performs an action on the dataset at the time of classification. K-NN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is the most similar to the new data.

The K-NN works as follows: The distance between the test data and all the training samples are calculated. The distance may be calculated by any standard means. Euclidean distance is usually used. The K nearest neighbor may be considered if the distance of the training samples to the test samples is less than or equal to K^{th} smallest distance. The quality of the predictions relies on the distance measure. The K-NN algorithm is suitable for applications that have sufficient domain knowledge.

B. Naive Bayes

Naive Bayes is one of the probabilistic learning method proposed by respected British scientist Thomas Bayes. Naive Bayes often works much better in many complex real-world situations than might be expected [13]. It is a popular model in Machine Learning applications because of its simplicity in allowing all attributes to contribute to the final decision equally. This simplicity is equivalent to computational efficiency, which makes the Naive Bayes technique attractive and suitable for various fields. The main element of Naive Bayes Classifier are the prior, posterior and conditional probability.

The Bayes Theorem formula can be seen in equation 2:

$$P(Q/X) = \frac{P(X/Q)P(Q)}{P(X)} \tag{2}$$

where X = data with unspecified class, Q = hypothesis, X is the specified class, $P\ (Q\ /\ X)$ = probability with hypothesis Q refers to X, $P\ (Q)$ = probability with hypothesis Q (previous probability), $P\ (X\ /\ Q)$ = probability X with hypothesis Q and $P\ (X)$ = probability X.

C. Decision Tree

A decision tree is a supervised classification method. A decision tree is a classifier P: $a \rightarrow b$, which classifies the label associated with an instance a by traveling from a root node of a tree to a leaf [12]. For simplicity we focus on the binary classification setting, namely, $b = \{0, 1\}$, but decision trees can be applied for other prediction problems as well. At each node on the root-to-leaf path, the successor child is chosen on the basis of a splitting of the input space. The splitting is usually based on one of the features of a or on a predefined set of splitting rules. A leaf contains a specific label.

D. Random Forest

Leo Breiman designs a Random Forest, which is a collection of classification or regression trees generated from a random selection of training data samples. A group of decision trees consists of a random forest classifier, where the entire tree is run by applying algorithm A to the training set S and an additional random vector θ of a distribution [12]. The prediction of the random forest is obtained by a majority vote over the predictions of the individual trees.

To specify a particular random forest, we need to define the algorithm A and the distribution over θ. There are many ways to define Algorithm A and here we describe one particular option. We generate θ as follows. First, we take a random subsample from S with replacements; namely, we sample a new training set S_0 of size M_0 using the uniform distribution over S. Second, we construct a sequence I_1, I_2, ..., where each I is a subset of [d] of size k, which is generated by sampling uniformly at random elements from [d]. All these random variables are form the vector θ. Then, the algorithm A grows a decision tree (e.g., using the ID3 algorithm) based on the sample S', where at each splitting stage of the algorithm, the algorithm is restricted for choosing a feature that maximizes Gain from the set I_t. Intuitively, if k is small, this constriction may prevent overfitting.

E. Support Vector Machines

Supporting vector machines (SVM) [14] are the most important development of machine learning algorithms based on kernel methods. Prediction, classification or other works can be done using support vector machines. SVM sets up a hyperplane or set of hyperplanes in high or infinite dimensional space. Intuitively, it is clear that the hyperplane provides good separation, which has the greatest distance of points near the training data of any class (the so-called functional margin), because the greater the generalization error of the classifier, the lower the overall margin. The support-vector network implements the following idea: it maps the input vectors into some high dimensional feature space Z through some non-linear mapping chosen a priori. In this space, a linear decision surface is constructed with special properties that ensure high generalization ability of the network.

F. XGBoost

XGBoost (eXtreme Gradient Boosting) is an open-source library that provides an efficient and effective implementation of the gradient boosting algorithm. The gradient amplification decision tree algorithm is used by the XGBoost library [15]. This algorithm is also known as gradient boosting, multiple additive regression trees, stochastic gradient boosting, or gradient boosting machines. Boosting is an ensemble technique where existing models make new models by adding to correct the errors. Until no further improvements can be made, models are added sequentially.

AdaBoost [15] is a popular example that weights data points that are hard to predict. Gradient boosting is an approach which create new models for predicting the residuals or errors of prior models and then added together to make the final prediction. It is called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models. Both regression and classification predictive modeling problems are supported by this approach.

VI. Experiment

We designed experiments to confirm the effectiveness of the proposed two-step approach.

A. Corpus Development

We collected 10,853 real questions that the farmers asked from six Myanmar Agriculture applications: Htwet Toe [20], Green Way [21], Hnit Shay [22], Plant Protection [23], Golden Paddy [24] and DOA [25]. Word segmentation, tagging with defined tags for each word, labelling with defined classes for each question and error checking were done manually. We used the tagsets and classes that we mentioned in Section III.

B. Evaluation

The hold out evaluation method is used for measuring the performance by utilizing the following equations:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TN}{TP + FN} \tag{4}$$

$$F1\,Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

where TP is the total number of true records for each class that are correctly classified, FP is the total number of true records for each class that are incorrectly classified, TN is the total number of false records for each class that are correctly classified and FN is the total number of false records for each class that are incorrectly classified.

We used SCTK (NIST Scoring Toolkit) version 2.4.10 [19] to dynamically construct programming matches between reference tag strings and hypothesis strings and calculate the word error rate (WER). The formula can be formulated as follows:

$$F1\,Score = \frac{(I + D + S) * 100}{N} \tag{6}$$

where S is the total number of replacements, D is the total number of deletions, I is the the total number of insertions, C is the total number of correct words and N is the total number of words in the sentence N=(S+D+C).

C. Implementation

We implemented the function tagging experiment by using the following open source Taggers:

1) CRFSuite: We used the CRFsuite tool (version 0.12) [16], for training and testing CRFs models. Its speed is faster than other CRF toolkits.
2) Jitar (version 0.3.3): is a simple part-of-speech tagger based on a trigram Hidden Markov Model (HMM). Jitar is written in Java [17] and thus easy to use in other Java programs, or languages that run on the JVM.
3) RDRPOSTagger (Version 1.2.3): is a rule-based Part-of-Speech tagging toolkit [9] [10]. It is a powerful, simple and non-linguistic toolkit.

For question classification, we also carried out our experiment using open source python libraries [18] such as KneighborsClassifier, Naive Bayes (MultinomialNB), DecisionTreeClassifier, RandomForestClassifier, LinearSVC (Linear Support Vector Classification) and XGBClassifier.

VII. Result and Discussion

A. Function Tagging

We conduct experiments of function tagging with three tagging approaches. We splitted the dataset on the ratio of 90% for training and 10% for testing purposes. Table IV shows the tagging result. As a result in Table IV, we got a high accuracy of over 90% for each task. RDR gives the highest F1 score of 0.96. These results can be used for the succeeding task of question classification because the accuracies of three tagging approaches are so good.

TABLE IV: F1-score of Three Tagging Approaches

Tagging Approaches	**CRFs**	**HMM**	**RDR**
F1-score	0.94	0.95	**0.96**

B. Question Classification

We conduct experiments to examine whether the function tagging is effective for question classification or not. For that purpose, we construct the following three classification models:

1) Classification without Tags: Classify the input questions without tagging process as shown in Figure 1.
2) Classification with Predicted Function Tags: Classify the questions tagged by the proposed system such as CRFs, HMM and RDR as shown in Figure 2.
3) Classification with Manual Tags: Classify the questions tagged by manually.

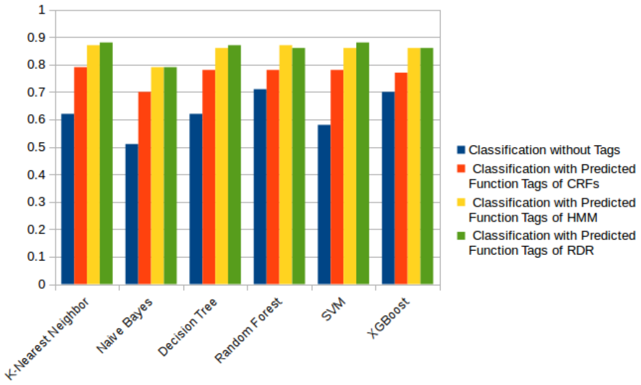

Fig. 5: F1-score of Classification on untagged Question and on tagged Question of CRFs, HMM and RDR

Figure 5 shows that the F1-score of classification on untagged Question and on tagged Question of CRFs, HMM and RDR tagging approaches. On the other hand, this figure compares the results of one-step in Figure 1 and two-step in Figure 2. In this figure, Classification without Tags is one-step approach's classification results and the others are two-step approach's results using K-Nearest Neighbor, Naive Bayes, Decision Tree, Random Forest, SVM and XGBoost. In one-step (classification on untagged question) question classification, the F1-score of all proposed classifier models is decreased a lot than two-step (classification on tagged question). Therefore, we found that function tagging is effective on question classification.

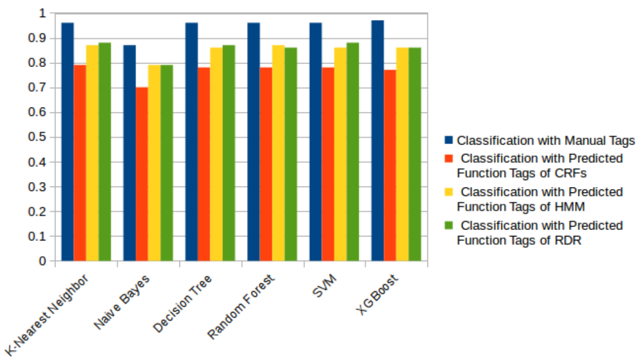

Fig. 6: F1-score of Classification on manual tagged Question and on tagged Question of CRFs, HMM and RDR

Figure 6 also shows that the F1-score of question classification on auto-tagging and manual tagging. In this figure, we compare the result that is classified by using manually tagged questions and the results that is classified by using auto-tagged questions of CRFs, HMM and RDR. We found that the F1 score of all classifier models with manual tags of question is so dramatically increased than auto-tagged questions of CRFs, HMM and RDR. In auto-tagged questions, the F1 score of all proposed models that is classified by using the tags defined by CRFs is slightly decreased than by using the tags defined by HMM and RDR. The tagging result of CRFs is also slightly decreased than HMM and RDR as described in Table IV. It is clear that, as the tagging accuracy of CRFs decreases, classification accuracy is also decreased. Therefore, we can claim that if the function tagging is accurate, question classification is more accurate. According to this result, we can still expect some more improvements of performance, by boosting the accuracy of function tagging.

C. Error Analysis

The WER calculations for the CRFs, HMM and RDR outputs were performed using the SCLITE program from the SCTK toolkit [19]. The lower the WER, the better the result. The WER results of the three tagging approaches can be seen in Table V. In this result, the RDR is slightly lower than the CRFs and HMM approaches.

TABLE V: WER of CRFs, HMM and RDR Tagging Approaches

Model	**CRFs**	**HMM**	**RDR**
WER	6.3(%)	5.2(%)	**4.4(%)**

Error analysis of three tagging approaches have also been done with the help of confusion matrices produced by the SCLITE program. We studied all confusion pairs for tagging approaches (including CRFs, HMM and RDR) and found that confusion pairs are different according to the tagging methodologies. After analyzing the confusion pairs of CRFs, HMM and RDR, the confusion pairs are most commonly caused by unknown words (Out-Of-Vocabulary) that are not in training. The current tagger fails to label a tag when the system meets an unknown word, since this word is not trained by the system. Handling unknown words in the tagger is also an essential need. For examples, the top four confusion pairs of three tagging approaches can be seen in Table VI, Table VII and Table VIII.

TABLE VI: Some Confusion Pairs of CRFs Tagging Approach

Frequency	REF ==> HYP
2	ခြောက်ပြီးကြွေ/bacd ==> ခြောက်ပြီးကြွေ/nolabel
2	ဂန္ဓာမာ/flo ==> ဂန္ဓာမာ/nolabel
2	ဆေးမြင်းခွာ/veg ==> ဆေးမြင်းခွာ/nolabel
2	ထိရောက်အောင်/nolabel ==> ထိရောက်အောင်/agrk

In tagging experiment, F1-score of RDR is the highest. So, the classification results of K-NN, Decision Tree, SVM and XGBoost on auto-tagging with RDR are better than

TABLE VII: Some Confusion Pairs of HMM Tagging Approach

Frequency	REF ==> HYP
2	ကာကွယ်ကုသတဲ့/nolabel ==> ကာကွယ်ကုသတဲ့/bacd
2	ကျောက်ရောဂါ/bacd ==> ကျောက်ရောဂါ/nolabel
2	ခြောက်ပြီးကြွေ/bacd ==> ခြောက်ပြီးကြွေ/nolabel
2	ဖြုတ်စိမ်းဆေး/bacd ==> ဖြုတ်စိမ်းဆေး/nolabel

TABLE VIII: Some Confusion Pairs of RDR Tagging Approach

Frequency	REF ==> HYP
2	ခြောက်ပြီးကြွေ/bacd ==> ခြောက်ပြီကြွေ/nolabel
2	ဂန္ဓာမာ/flo ==> ဂန္ဓာမာ/nolabel
2	ဆေးမြင်းခွာ/veg ==> ဆေးမြင်းခွာ/nolabel
2	ရောဂါ/bacd ==> ရောဂါ/nolabel

other two tagging approaches. The F1-score of CRFs is the lowest in function tagging. Due to the lowest auto-tagging result of the CRFs, the proposed all classifier models on auto-tagging with CRFs gave the lowest clasification results. Therefore, accurate question classification can be very hard without accurate tagging process.

On the other hand, someone may consider that the type of questions can be classified by "one-step" approach without function tagging. Therefore, the proposed system was compared with this "one-step" approach in the initial experiment. We found that our two-step approach is better than "one-step" as this experiment.

VIII. Conclusion

In this paper, a model was proposed for identifying the type of questions. In our model, the input question is first tagged and then these tags are used for question classification. RDR gave maximum F1-score of 0.96 in function tagging. K-Nearest Neighbor and SVM also gave the maximum F1-score of 0.88 in final question classification by using the tags defined by RDR. We experimented the effectiveness of function tagging in the question classification. In function tagging, handling the unknown word is necessary. To handle this error, we need to increase the training data. In Burmese question classification, function tagging is quite important and we studied well known machine learning approaches. As future work, we plan to classify with multiple classification approaches based on neural networks such as CNN and RNN. We also plan to check errors of manual segmentation, tagging of the whole corpus, reevaluate with cross validation and increase corpus size.

References

[1] Prashant Y Niranjan, Vijay S Rajpurohit and Rasika Malgi, "A Survey on Chat-Bot system for Agriculture Domain," International Conference on Advances in Information Technology, 2019, pp. 99–103.

[2] Hasangi Kahaduwa, Dilshan Pathirana, Pathum Liyana Arachchi, Vishma Dias, Surangika Ranathunga and Upali Kohomban, "Question Answering System for the Travel Domain," Moratuwa Engineering Research Conference (MERCon), 2017, pp. 449–450.

[3] Joseph Lilleberg Yun Zhu and Yanqing Zhang,Eiichiro Sumita and Chiori Hori, "Support Vector Machines and Word2vec for Text Classification with Semantic Features," International Conference on Cognitive Inlormatics & Cognitive Computing, 2015, pp. 136-139.

[4] Manmita Devi and Mohit Dua, "ADANS: An Agriculture Domain Question Answering System using Ontologies," International Conference on Computing, Communication and Automation, 2017, pp. 264–267.

[5] Ji Ho Park and Pascale Fung, "One-step and Two-step Classification for Abusive Language Detection on Twitter," Proceedings of the First Workshop on Abusive Language Online, 2017, pp. 41-45.

[6] K. S. Kalaivani, R. Felicia Grace, M. Aarthi and M. Boobeash, "Classification of Sentiment Reviews using POS based Machine Learning Approach," International Journal of Engineering Research and Technology (IJERT), 2018, pp. 1-5.

[7] John Lafferty, Andrew McCallum and Fernando CN Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001), pp. 282-289.

[8] Virtusa, Doctor Danister De Silva Mawatha, Colombo and Sri Lanka, "Hidden Markov Model Based Part of Speech Tagger for Sinhala Language," International Journal on Natural Language Computing, 2014, pp. 9-16.

[9] Nguyen, Dat Quoc, Dai Quoc Nguyen, Dang Duc Pham and Son Bao Pham, "RDRPOSTagger: A Ripple Down Rules-based Part-Of-Speech Tagger," In Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics, 2014, Kyoto, Japan. pp. 17-20.

[10] Nguyen, Dat Quoc, Dai Quoc Nguyen, Dang Duc Pham and Son Bao Pham, "A Robust Transformation-based Learning Approach using ripple down rules for part-of-speech tagging," AI Communications, pp 409–422.

[11] Akiko Aizawa, "An information-theoretic perspective of tf–idf measures," Information Processing and Management 39, 2003, pp. 49-54.

[12] Shai Shalev-Shwartz and Shai Ben-David, "Understanding Machine Learning: From Theory to Algorithms," version 1, 2014, pp 250-256, 258-267.

[13] Aji Prasetya Wibawa, "Naive Bayes Classifier for Journal Quartile Classification," International Journal of Recent Contributions from Engineering, Science and IT (iJES), 2019, pp. 94-96.

[14] Aji Prasetya Wibawa, "Support-Vector Networks," Kluwer Academic Publishers, Boston Manufactured in The Netherlands, 1995, pp. 273-297.

[15] Jason Brownlee, "Gradient Boosted Tree with XGBoost and scikit-learn," Edition v1 10, pp. 2-22.

[16] Okazaki and Naoaki, "CRFsuite: a fast implementation of Conditional Random Fields (CRFs), 2016.

[17] de Kok and Daniël, "Jitar: A simple Trigram HMM part-of-speech tagger," 2014.

[18] Fabian Pedregosa, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research 12, 2011, pp. 2825-2830.

[19] "Speech recognition scoring toolkit (sctk)," The National Institute of Standards and Teechnology (NIST), version 2.4.10, 2015.

[20] "Htwet Toe," January, 2018. [Online]. Available: https://htwettoe.com.

[21] "Green Way," May, 2011. [Online]. Available: https://www.mmgreenovator.com.

[22] "Hnit Shay," October, 2018. [Offline]. Available: https://play.google.com/store/apps/details?id=com.mmhash.perennialcropsdivision&hl=en_US&gl=US.

[23] "Plant Protection," February, 2020. [Offline]. Available: https://play.google.com/store/apps/details?id=com.averta.plantprotection&hl=en&gl=US.

[24] 'Golden Paddy," December, 2016. [Online]. Available: https://www.shwetheenhan.com.

[25] "DOA," [Online]. Available: http://www.doa.gov.mm/doa/.

Cho Zin Oo got her master degree of Computer Science from University of Computer Studies, Mandalay, Myanmar. She is a tutor of Faculty of Information Science (FIS), University of Computer Studies, Hinthada and also a Ph.D candidate at University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar. She is currently pursuing her Ph.D. studies in Question Classification for Automatic Question-Answering in Agriculture Domain.

Ye Kyaw Thu is a Visiting Professor of Language & Semantic Technology Research Team (LST), Artificial Intelligence Research Unit (AINRU), National Electronic & Computer Technology Center (NECTEC), Thailand and Head of NLP Research Lab., University of Technology Yatanarpon Cyber City (UTYCC), Pyin Oo Lwin, Myanmar. He is also a founder of Language Understanding Lab., Myanmar and a Visiting Researcher of Language. He is actively co-supervising/supervising undergrad, masters' and doctoral students of several universities including KMITL, SIIT, UCSM, UCSY, UTYCC and YTU.

Hlaing Myat Nwe is a PhD candidate of Sirindhorn International Institute of Technology, Thammasat University, Thailand. A native of Myanmar, she holds a master degree of Information Science and Technology, and a bachelor degree of Information Science and Technology from University of Technology (Yatanarpon Cyber City), Myanmar. Her research interests include human-computer interaction, natural language processing, audio signal and image processing. She has been working to find efficient and user-friendly text input interfaces and video translation system for Myanmar Sign Language and Myanmar SignWriting. She is also a supervising lab member from NLP-Lab, UTYCC.

Hnin Aye Thant She is currently working as a Professor and Head of Department of Information Science at the University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin Township, Mandalay Division, Myanmar. She-got Ph.D (IT) Degree from University of Computer Studies, Yangon, Myanmar in 2005. The current responsibilities are managing professional teachers, doing instructional designer of e-learning content development and teaching. She has 14 years teaching experiences in Information Technology specialized in Programming Languages (C,C++, Java and Assembly), Data Structure, Design and Analysis of Algorithms/Parallel Algorithms, Database Management System, Web Application Development, Operating System, Data Mining and Natural Language Processing. She is a member of research group in "Neural Network Machine Translation between Myanmar Sign Language to Myanmar Written Text" and Myanmar NLP Lab in UTYCC. She is also a Master Instructor and Coaching Expert of USAID COMET Mekong Learning Center. So, she has trained 190 Instructors from ten Technological Universities, twelve Computer Universities and UTYCC for Professional Development course to transform teacher-centered approach to learner-centered approach. This model is to reduce the skills gap between Universities and Industries and to fulfill the students' work-readiness skills.