Integrating Genetic Algorithms with Conditional Random Fields to Enhance Question Informer Prediction

Min-Yuh Day [a, b], Chun-Hung Lu [a, b], Chorng-Shyong Ong [b], Shih-Hung Wu [c], and Wen-Lian Hsu [a,*], *Fellow, IEEE*

[a] *Institute of Information Science, Academia Sinica, Taipei 115, Taiwan*

[b] *Department of Information Management , National Taiwan University, Taipei, 106, Taiwan*

[c] *Dept. of Computer Science and Information Engineering, Chaoyang Univ. of Technology, Taiwan*

{myday,enrico,hsu}@iis.sinica.edu.tw; ongcs@im.ntu.edu.tw; shwu@cyut.edu.tw

Abstract

Question informers play an important role in enhancing question classification for factual question answering. Previous works have used conditional random fields (CRFs) to identify question informer spans. However, in CRF-based models, the selection of a feature subset is a key issue in improving the accuracy of question informer prediction. In this paper, we propose a hybrid approach that integrates Genetic Algorithms (GAs) with CRF to optimize feature subset selection in CRF-based question informer prediction models. The experimental results show that the proposed hybrid GA-CRF model improves the accuracy of question informer prediction of traditional CRF models.

1. Introduction

Question informers play an important role in enhancing question classification for factual question answering [3]. Krishnan et al. [3] introduced the notion of the answer type informer span of a question for question classification and showed that human-annotated informer spans lead to large improvements in the accuracy of question classification. They defined that choosing a minimal, appropriate contiguous span of a question token, or tokens, as the informer span of a question, which is adequate for question classification. For example, in the question: “What is the biggest city in the United State?” the question informer is “city”. Thus “city” is the most important clue in the question for question classification. Krishnan et al. reported that perfect knowledge of informer spans can enhance the predictive accuracy from 79.4% to 88% using linear Support Vector Machines (SVMs) on standard benchmarks of question classification.

Question informers are only useful if question informer spans can be identified automatically. Previous works have used Conditional Random Fields (CRFs) to identify question informer spans. By using a parse of the question, Krishnan et al. [3] derived a set of multi-resolution features to train a CRF model and achieved 85%-87% accuracy for question informer prediction.

Krishnan et al. [3] also showed that the effect of the features chosen by a CRF model varies significantly depending on the accuracy of the CRF model. In a machine learning approach, feature selection is an optimization problem that involves choosing an appropriate feature subset. In CRF-based models, selection of the feature subset is a key issue in improving the accuracy of question informer prediction. Genetic Algorithms (GAs) [2] have been widely used in feature selection in machine learning [10].

In this paper, we propose a hybrid approach that integrates GA with CRF to optimize feature subset selection in CRF-based question informer prediction models.

The remainder of this paper is organized as follows. Section 2 describes the background to question informers and previous works. In section 3, we propose the hybrid GA-CRF approach for question informer prediction. Section 4 discusses the experiment and the test bed, and Section 5 contains the experimental results. Finally, in Section 6 we present our conclusions and indicate some future research directions.

2. Research Background

2.1. Conditional Random Fields (CRFs)

Lafferty et al. [4] proposed using Conditional Random Fields (CRFs), a framework for building probabilistic models, to segment and label sequence data. A CRF models *Pr(y|x)* using a Markov random field, with nodes corresponding to elements of the structured object y, and potential functions that are conditional on features of x. Learning is performed by setting parameters to maximize the likelihood of a set of *(x,y)* pairs given as training data [8].

CRFs are widely used for sequential learning problems like NP chunking, POS tagging, and name entity recognition (NER). Recent works [1, 4, 8] have shown that CRFs have a consistent advantage over traditional Hidden Markov Models (HMMs) and Maximum Entropy Markov Models (MEMMs) [6] in the face of many redundant features. Krishnan et al. [3] reported that they achieved 85%-87% accuracy of question informer prediction by using CRF model with a set of features.

CRF++, which is developed by Taku Kudo, is a simple, customizable, and open source implementation of CRFs for segmenting and labeling sequenced data (CRF++ is available at: http://chasen.org/~taku/software/CRF++/). It was designed for generic purposes and can be applied to a variety of NLP tasks, such as Named Entity Recognition, Information Extraction, and Text Chunking. The benefit of using CRF++ is that it enables us to redefine feature sets and specify the feature templates in a flexible way.

2.2. Genetic Algorithm (GA)

Genetic Algorithms (GAs) are a class of heuristic search methods and computational models of adaptation and evolution based on the mechanics of natural selection and genetics [2]. GAs have been widely used for feature selection in machine learning [10] methods, such as SVM. Feature selection is an optimization problem that involves the process of picking a subset of features that are relevant to the target concept and removing irrelevant or redundant features. This is an important factor that determines the performance of the machine learning models.

3. The Hybrid GA-CRF Model

In this paper, we propose integrating the GA architecture with CRF to optimize feature selection for CRF-based question informer prediction. Fig 1 shows the architecture of the proposed hybrid GA-CRF approach for question informer prediction. There are two phases in the architecture. The first is the GA-CRF learning phase with a training dataset, while the second is the CRF test phase with a test dataset.

The application of GA to CRF-based question informer prediction comprises the following steps.

1) Encoding a feature subset of CRF with the structure of chromosomes: To apply GA to the search for the optimal feature subsets of CRF, the subsets must be encoded on a chromosome in the form of binary strings. The gene structure of the chromosomes for feature subset selection is presented in Figure 2.

The value of the codes for feature subset selection is set to a one-bit digit '0' or '1', where '0' means the corresponding feature is not selected, and '1' means that it is selected. The length of each chromosome is n bits, where n is the number of features. We use f_{i-2}, f_{i-1}, f_{i+0}, f_{i+1}, f_{i+2} to represent the sliding windows of each feature. Figure 3 shows an example of feature subset encoding for GA.

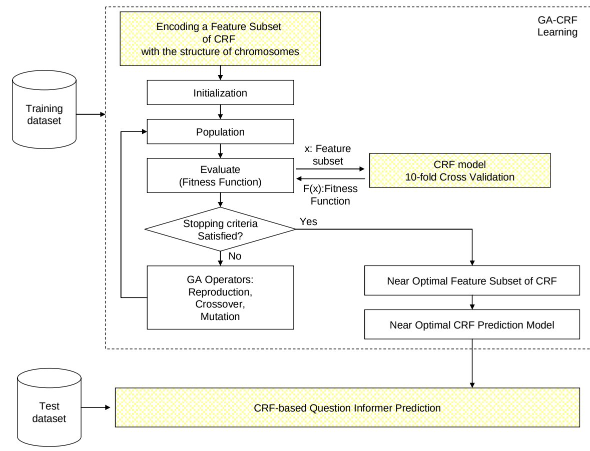


Figure 1. The architecture of proposed hybrid GA-CRF approach for question informer prediction

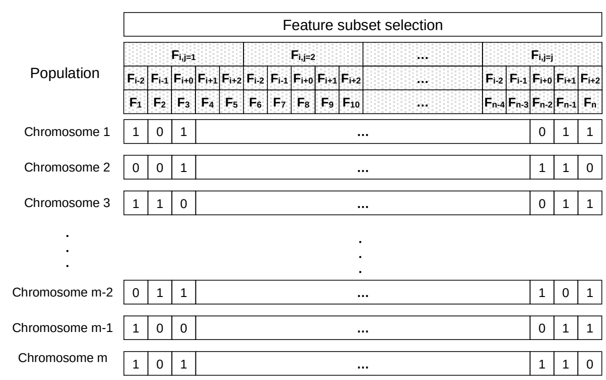
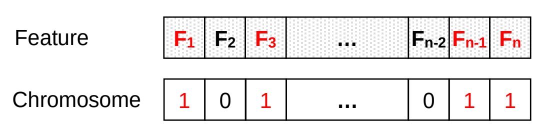


Figure 2. Gene structure of chromosomes for a feature subset



Feature subset = $\{F_1, F_3, \ldots, F_{n-1}, F_n\}$

Figure 3. An example of feature subset encoding for GA

2) Initialization: In this step, we generate the initial population, which is initialed into random values before the search process.

3) Population: The population is a set of seed chromosomes used to find the optimal feature subsets.

4) Evaluate (Fitness Function): In this step, we calculate the fitness score of each chromosome. In addition, the population is searched to find the encoded chromosome that maximizes the specific fitness function. The values of the fitness functions for the items in the evaluation set are calculated and used to determine the suitability of each chromosome.

5) CRF model 10-fold Cross validation: In this procedure, the feature subsets derived by the previous procedure are applied to the CRF module. The fitness function is determined by the F-score of 10-fold cross validation of the CRF model. We use 10-fold cross validation on the training dataset of the CRF model as the fitness function of each chromosome to avoid over-fitting on the test dataset.

6) Stopping criteria satisfied? If the stopping criteria are satisfied, the best chromosome and a near optimal feature subset of CRF model is obtained; otherwise, apply GA operators and produce a new generation.

7) Apply GA operators and produce a new generation: In this procedure, we use three GA operators, namely, reproduction, crossover, and mutation to produce a new generation.

8) Apply the selected feature subsets to the CRF test dataset: After the GA-CRF learning process, we can obtain a near optimal feature subset of CRF. We then train the whole training set on that feature subset to obtain a near optimal CRF prediction model, which we use to test the test dataset for CRF-based question informer prediction.

4. Experimental Design

4.1. Data set

We use the UIUC QC dataset from Li and Roth [5] and the corresponding question informer dataset from Krishnan et al. [3]. There are 5,500 training questions, 500 test questions, and corresponding question informers[1]. Li and Roth used supervised learning for question classification of the UIUC QC data set, which is now the standard dataset for question classification [5, 9, 11]. It has 6 coarse and 50 fine answer types in a two level taxonomy, together with 5,500 training and 500 test questions. Krishnan et al. [3] reported that they had two volunteers to tag 6,000 UIUC questions with informer spans, which they call human-annotated "perfect" informer spans.

4.2. Features of Question Informer

We adopt the 3-state transition model suggested by Krishnan et al [3] and follow the "begin/in/out" (BIO) model proposed by Ramshaw and Marcus [7] to tag question informers. In our dataset for the CRF model, "O-QIF0" indicates outside and before a question informer, "B-QIF1" indicates the start of a question informer, while "O-QIF2" indicates outside and after a question informer.

Table 1 shows 21 basic feature candidates for question informer prediction. Features 1 to 4 are word, POS, heuristic informer, and question segmentation information, respectively. Features 5 to 11 are parser information about the question. We use the OpenNLP Parser (available at: http://opennlp.sourceforge.net/) to parse a question and translate the parse tree into a two-dimensional matrix with 6 levels as features from the parse of a question. Although the parse tree can be arbitrarily deep, Krishnan et al. [3] indicate that using features with 2 levels is adequate. Features 12 to 18 are derived from heuristics that are suggested by Krishnan et al. We add Features 19 to 21, namely, question wh-word (6W1H1O: who, what, when, where, which, why, how, and others), question length, and token position.

In this study, we regard each feature candidate as a gene, and treat the corresponding F-score as the performance value of the feature (gene) for the CRF model. Table 1 shows that word, POS, parser level 1, and parser level 2 have better performance for the CRF model. For example, the F-score of a single feature used with the "word" feature is 58.35%, while using the "parser level 2" feature solely achieves a score of 48.13%. The experiment result shows that each feature candidate has a different effect on the performance of CRF-based question informer prediction.

Figure 4 shows an example of a feature with sliding windows for a CRF model. For example, "city" is the feature f_{ij} for x_i, where $i = 4$ and $j=0$. Given that x_4 = "city", the label of prediction y_4 = "B-QIF1".

Figure 5 shows an example of feature generation and a feature template for CRF++. For example, we can specify the feature "city" in feature $f_{0,0}$ as the feature template "U02:%[0,0]", and the previous feature "oldest" in feature $f_{-1,0}$ as the feature template "U01:%[-1,0]".

[1] Li and Roth (2002), UIUC QC Datasets: http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/QC/
Vijay Krishnan, Sujatha Das and Soumen Chakrabarti (2005), UIUC Informers Datasets : http://hake.stanford.edu/~kvijay/UIUC_Informers/

Question: What is the oldest city in the United States?
Data Format for CRF++

	Features f_{ij} for x_i																					
j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
i	x_i	POS	HQI	Token	ParL0	ParL1	ParL2	ParL3	ParL4	ParL5	ParL6	IsTag	IsNum	IsPrevTag	IsNextTag	IsEdge	IsBegin	IsEnd	Wh-word	L	P	y_i
0	What	WP	city	What	What	WP_1	WHNP_1	Null_1	Null_1	Null_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag1	IsNextTag0	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	1	O-QIF0
1	is	VBZ	city	is	is	VBZ_1	Null_1	Null_1	VP_1	SQ_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag1	IsNextTag0	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	2	O-QIF0
2	the	DT	city	the	the	DT_1	NP_1	Null_1	VP_1	SQ_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag1	IsNextTag0	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	3	O-QIF0
3	oldest	JJS	city	oldest	oldest	JJS_1	NP_1	Null_1	VP_1	SQ_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag1	IsNextTag0	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	4	O-QIF0
4	city	NN	city	city	city	NN_1	NP_1	Null_1	VP_1	SQ_1	SBARQ_1	IsTag1	IsNum1	IsPrevTag0	IsNextTag0	IsEdge1	IsBegin1	IsEnd1	Wh_what	10	5	B-QIF1
5	in	IN	city	in	in	IN_1	Null_2	PP_1	VP_1	SQ_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag0	IsNextTag1	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	6	O-QIF2
6	the	DT	city	the	the	DT_2	NP_2	PP_1	VP_1	SQ_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag0	IsNextTag1	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	7	O-QIF2
7	United	NNP	city	United	United	NNP_1	NP_2	PP_1	VP_1	SQ_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag0	IsNextTag1	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	8	O-QIF2
8	States	NNS	city	States	States	NNS_1	NP_2	PP_1	VP_1	SQ_1	SBARQ_1	IsTag0	IsNum0	IsPrevTag0	IsNextTag1	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	9	O-QIF2
9	?	.	city	?	?	._1	Null_3	Null_2	Null_2	Null_2	SBARQ_1	IsTag0	IsNum0	IsPrevTag0	IsNextTag1	IsEdge0	IsBegin0	IsEnd0	Wh_what	10	10	O-QIF2

Figure 6. An example of the data format of a question informer and the corresponding features for a CRF model

We generate the 21 feature candidates in a two-dimensional matrix and add the question informer tag (O-QIF0, B-QIF1, O-QIF2) for each question sentence in the last column. Figure 6 shows the data format of a question informer and the corresponding features candidates in a CRF model for the question "What is the oldest city in the United States?".

Table 1. Features for question informer prediction

ID	Feature name	Description	Feature Template for CRF ++	F-score	Feature Rank
1	Word	Word	U01:%x[0,0]	58.35	1
2	POS	POS	U01:%x[0,1]	48.29	6
3	HQI	Heuristic Informer	U01:%x[0,2]	52.21	4
4	Token	Token	U01:%x[0,3]	58.35	2
5	ParserL0	Parser Level 0	U01:%x[0,4]	58.35	3
6	ParserL1	Parser Level 1	U01:%x[0,5]	50.98	5
7	ParserL2	Parser Level 2	U01:%x[0,6]	48.13	7
8	ParserL3	Parser Level 3	U01:%x[0,7]	37.76	9
9	ParserL4	Parser Level 4	U01:%x[0,8]	38.45	8
10	ParserL5	Parser Level 5	U01:%x[0,9]	21.45	17
11	ParserL6	Parser Level 6	U01:%x[0,10]	22.43	13
12	IsTag	Is Tag	U01:%x[0,11]	21.57	15
13	IsNum	Is Number	U01:%x[0,12]	21.57	16
14	IsPrevTag	Is Previous Tag	U01:%x[0,13]	21.21	18
15	IsNextTag	Is Next Tag	U01:%x[0,14]	28.75	11
16	IsEdge	Is Edge	U01:%x[0,15]	21.58	14
17	IsBegin	Is Begin	U01:%x[0,16]	15.45	20
18	IsEnd	Is End	U01:%x[0,17]	28.26	12
19	Wh-word	Question Wh-word (6W1H1O)	U01:%x[0,18]	30.17	10
20	Length	Question Length	U01:%x[0,19]	20.93	19
21	Position	Token Position	U01:%x[0,20]	13.17	21

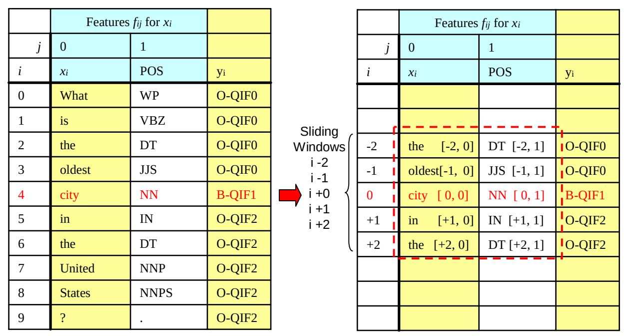


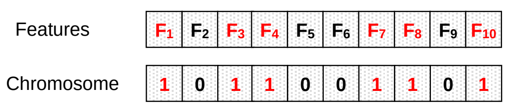
Figure 4. An example of feature with sliding windows for CRF

Feature	Features	Feature Template	Feature ID
the	f-2,0	U00:%x[-2,0]	F1
oldest	f-1,0	U01:%x[-1,0]	F2
city	f0,0	U02:%x[0,0]	F3
in	f+1,0	U03:%x[+1,0]	F4
the	f+2,0	U04:%x[+2,0]	F5
DT	f-2,1	U05%x[-2,1]	F6
JJS	f-1,1	U06:%x[-1,1]	F7
NN	f0,1	U07:%x[0,1]	F8
IN	f+1,1	U08:%x[+1,1]	F9
DT	f+2,1	U09:%x[+2,1]	F10

Figure 5. An example of feature generation and a feature template for CRF++

4.3. Encoding a feature subset with the structure of chromosomes for GA

We encode all 21 feature candidates and sliding windows with the structure of chromosomes to form the feature subset for GA. The candidates corresponding sliding window size is 5 (*w*= *-2, -1, 0, +1, +2*), We use f_{i-2}, f_{i-1}, f_{i+0}, f_{i+1}, f_{i+2} to represent the windows of each feature candidate. Since there are 21 basic features and 5 sliding windows, we can generate 105 (21 basic features * 5 sliding windows) features (genes) for each chromosome. Figure 7 shows an example of encoding a feature subset with the structure of chromosomes for GA. For example, the chromosome "1 0 1 1 0 0 1 1 0 1" represents the selected feature subset is {F1, F3, F4, F7, F8 , F10}, and the corresponding feature template for CRF++ is { U00:%x[-2,0] U02:%x[0,0] U03:%x[+1,0] U06:%x[-1,1] U07:%x[0,1] U09:%x[+2,1]}.



Feature subset = {F_1, F_3, F_4, F_7, F_8 , F_{10}}

Features f_{ij} for x_i = Uid:*%x[i, j]*

Feature	Features	Feature Template	Feature ID
the	*f-2,0*	U00:%x[-2,0]	F1
oldest	*f-1,0*	U01:%x[-1,0]	F2
city	*f0,0*	U02:%x[0,0]	F3
in	*f+1,0*	U03:%x[+1,0]	F4
the	*f+2,0*	U04:%x[+2,0]	F5
DT	*f-2,1*	U05%x[-2,1]	F6
JJS	*f-1,1*	U06:%x[-1,1]	F7
NN	*f0,1*	U07:%x[0,1]	F8
IN	*f+1,1*	U08:%x[+1,1]	F9
DT	*f+2,1*	U09:%x[+2,1]	F10

Figure 7. Encoding a feature subset with the structure of chromosomes for GA. There are 105 feature subsets in total (21 basic features * 5 sliding windows)

4.4. Fitness function of GA

In the learning and validation phases, we use 10-fold cross validation with 5,500 UIUC training data to reduce the over-fitting problem, and use the selected near optimal feature subset for the CRF model.

5. Experimental Results

Figure 8 shows the experimental results of 10-fold cross validation on the training dataset and the corresponding performance on the test dataset using GA for feature subset selection of CRF-based question informer prediction. The F-score is approximately 95% for 10-fold cross validation on the training dataset (UIUC Q5500), and approximately 88% on the test dataset (UIUC Q500).

Figure 9 shows the experimental results of CRF-based question informer prediction using GA for feature subset selection for a population whose characteristics are: size = 40, crossover rate = 80%, and mutation rate = 10%. The F-score for question informer prediction is 93.87%. It should be noted that the fitness function is used to evaluate on the test dataset (UIUC Q500) with the training dataset (UIUC Q5500).

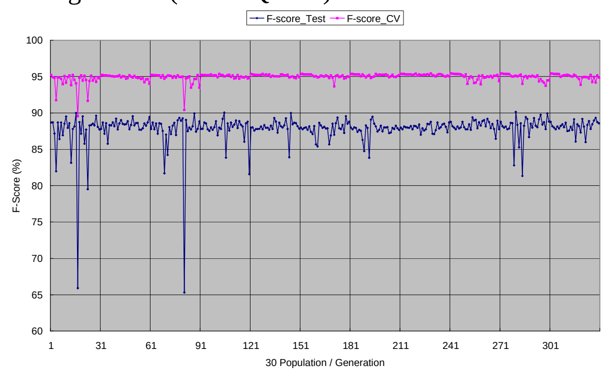


Figure8. 10-fold cross validation on the training dataset and their corresponding performance on the test dataset using GA for feature subset selection.

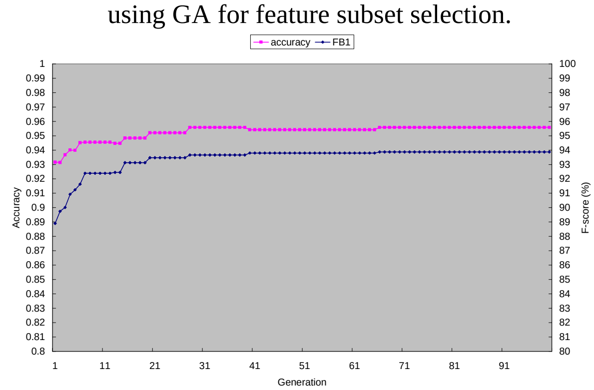


Figure 9. Experimental results of CRF-based question informer prediction using GA for feature subset selection. F Score: 93.87, Population: 40, Crossover: 80%, Mutation: 10%, Generation: 100

After 100 generations of GA, we obtain the near optimal chromosomes and their corresponding feature subset for CRF++. Figure 10 shows the near optimal chromosomes and the corresponding feature subset for the CRF model selected by GA. The experimental results show that we can improve the F-score of CRF-based question informer prediction from 88.9% to 93.87% using GA to reduce the number of features from 105 to a 40-feature subset.

The accuracy of our proposed GA-CRF model for UIUC dataset is 95.58% compared with 87% for the traditional CRF model reported by Krishnan et al. [3]. The experimental results show that our proposed hybrid

GA-CRF Model

Near Optimal Chromosome:

0 0 0 0 0 1 0 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 1 1 0 1 0 0 1 0 0 0 0 1 0 1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 1 1 0 0 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0 1 1 1 0 0 1

Near Optimal Feature Subsets for CRF model:

U001:%x[-2,1] U002:%x[0,1] U003:%x[1,1] U004:%x[-1,2] U005:%x[0,2] U006:%x[1,2] U007:%x[2,2] U008:%x[-2,3] U009:%x[-2,5] U010:%x[-1,5] U011:%x[-2,6] U012:%x[1,6] U013:%x[2,6] U014:%x[2,7] U015:%x[0,8] U016:%x[1,8] U017:%x[-2,9] U018:%x[1,9] U019:%x[1,10] U020:%x[-2,11] U021:%x[-2,12] U022:%x[0,12] U023:%x[0,13] U024:%x[2,13] U025:%x[-2,14] U026:%x[-1,14] U027:%x[2,14] U028:%x[0,16] U029:%x[1,16] U030:%x[2,16] U031:%x[-2,17] U032:%x[-1,17] U033:%x[0,17] U034:%x[1,17] U035:%x[-2,19] U036:%x[-1,19] U037:%x[2,19] U038:%x[-2,20] U039:%x[-1,20] U040:%x[2,20]

Question Informer Prediction	Accuracy	Recall	Precision	F-score
Traditional CRF Model (All features)	93.16%	94.33%	84.07%	88.90
GA-CRF Model (Near optimal feature subset)	95.58%	95.79%	92.04%	93.87

Figure 10. The near optimal chromosome and the corresponding feature subset for the CRF model selected by GA

GA-CRF model for question informer prediction outperforms the traditional CRF model.

6. Conclusions

We have proposed a hybrid approach that integrates Genetic Algorithm (GA) with Conditional Random Field (CRF) to optimize feature subset selection in a CRF-based model for question informer prediction. The experimental results show that the proposed hybrid GA-CRF model of question informer prediction improves the accuracy of the traditional CRF model. By using GA to optimize the selection of the feature subset in CRF-based question informer prediction, we can improve the F-score from 88.9% to 93.87%, and reduce the number of features from 105 to 40.

7. Acknowledgement

We would like to thank Vijay Krishnan for providing the question informer dataset and informative discussions. This research was supported in part by the thematic program of Academia Sinica under Grant AS 95ASIA02, the National Science Council under Grant NSC 95-2752-E-001-001-PAE, NSC 95-2416-H-002-047, and NSC94-2218-E-324-003.

8. References

[1] S. Fei and P. Fernando, "Shallow parsing with conditional random fields", Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (HLT-NAACL 2003), Edmonton, Canada, 2003.

[2] D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning: Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1989.

[3] V. Krishnan, S. Das, and S. Chakrabarti, "Enhanced Answer Type Inference from Questions using Sequential Models", in Proceedings of HLT/EMNLP. Vancouver, British Columbia, Canada, 2005, pp. 315-322.

[4] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", Proceedings of the Eighteenth International Conference on Machine Learning, 2001. pp. 282-289.

[5] X. Li and D. Roth, "Learning Question Classifiers", Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002), 2002.

[6] A. McCallum, D. Freitag, and F. Pereira, "Maximum Entropy Markov Models for Information Extraction and Segmentation", Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), 2000. pp. 591-598.

[7] L. A. Ramshaw and M. P. Marcus, "Text chunking using transformation-based learning", Proceedings of the Third ACL Workshop on Very Large Corpora, pp. 82-94, 1995.

[8] S. Sarawagi and W. W. Cohen, "Semi-markov conditional random fields for information extraction", Advances in Neural Information Processing Systems, vol. 17, pp. 1185–1192, 2005.

[9] J. Suzuki, H. Taira, Y. Sasaki, and E. Maeda, "Question Classification using HDAG Kernel", Workshop on Multilingual Summarization and Question Answering 2003 (post-conference workshop in conjunction with ACL-2003), 2003. pp. 61-68.

[10] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm", IEEE Intelligent Systems, vol. 13, pp. 44-49, 1998.

[11] D. Zhang and W. S. Lee, "Question classification using support vector machines", Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada, 2003. pp. 26-32.