Self-help Troubleshooting by Q-KE-CLD Based on a Fuzzy Bayes Model

Pilsung Choe[1], Mark R. Lehto[1], and Jan Allebach[2]

[1] School of Industrial Engineering, Purdue University, 1287 Grissom Hall, West Lafayette,
In 47907-1287, USA
[2] School of Electrical and Computer Engineering, Purdue University, 1285 Electrical
Engineering Building, West Lafayette, In 47907-1285, USA
{pchoe,lehto,allebach}@purdue.edu

Abstract. The previous study [9], [10] showed the fuzzy Bayes model successfully predicted print defects with a 50% hit rate at the first top prediction and an 80% hit rate within the top five predictions. However, the previous study was limited to English. In this study, Korean and English descriptions in predicting print defects by Korean subjects were evaluated based on fuzzy Bayes models. For the study, Korean descriptions were collected in Korea, and Bayes models were developed and evaluated. The result shows that Korean subjects much more accurately predicted print defects when they used Korean descriptions than English descriptions. Afterwards, English descriptions by US subjects will be collected, and both Korean and English lexicon data will be compared. Finally, the study will investigate a Korean-English cross language diagnosis (Q-KE-CLD) system to identify print defects based on the fuzzy Bayes model.

Keywords: fuzzy Bayes model, diagnosis, troubleshooting, Q-KE-CLD, cross language diagnosis, translation, CLIR, cross language information retrieval, parsing, lexical analysis, search.

1 Introduction

Due to the rapid spread of internet technology, troubleshooting using Web sites is becoming more popular because of its advantages in terms of economy and efficiency [2], [5], [6]. However, troubleshooting on Web sites written in English is often very difficult for users whose native language is not English, because of the difficulty of understanding English as well as the complexity of the troubleshooting process to navigate Web sites. In this respect, query-translation based cross language diagnosis (Q-CLD) using the fuzzy Bayes model can be a feasible solution for non-native English speaking users (NNEU) to troubleshoot on Web sites written in English because it avoids the process of navigating the complex Web sites written in English. The fuzzy Bayes model has been proven to be a good tool for identifying print quality issues [9], [10]. The result showed the performance was a 50% hit rate with the top one prediction, 70% with top three predictions, and 80% with top five predictions [9], [10]. However, the study was limited to English so that the result can not be directly

applied for NNEU to predict print defects, because the English descriptions used by NNEU may be different from those of native English speaking users. In addition, entering English queries is sometimes challenging for NNEU, although they use English in many situations such as reading, writing, and communicating. They often feel they could use their own languages as queries to find information they need, even though the information is written in English.

As a practical application, this study investigates a Korean-English cross language diagnosis (Q-KE-CLD) based on a fuzzy Bayes model to identify print quality defects. The original idea of cross language diagnosis (CLD) came from cross language information retrieval (CLIR), which is defined as finding relevant documents in a second language using a query expressed in a first language [1], [3], [11], [12]. Likewise, Q-KE-CLD is defined as finding relevant documents to diagnose problems in English using a query expressed in Korean. In order to investigate the Q-KE-CLD, four prediction cases are compared: (1) Korean predictions based on a Korean training set by Korean subjects; (2) English predictions based on an English training set by Korean subjects; (3) English predictions based on an English training set by US subjects; and (4) Korean predictions (which need translations to English) based on an English training set by Korean subjects. Case (1) or Case (3) is expected to provide an upper bound of the prediction accuracy, and case (2) is assumed to give a lower bound due to the deficiency of language ability of Korean subjects. In this aspect, case (4) is expected to have better prediction accuracy than case (3). The four cases are evaluated in terms of prediction accuracy, user performance, and user preference. In this paper, cases (1) and (2) are analyzed, and cases (3) and (4) will be conducted as a following study. Finally, the Q-KE-CLD system will be implemented and evaluated to determine whether the system can be a possible solution for localizing print quality troubleshooting Web sites [2], [5], [6].

2   Experiment for Model Development and Validation

In order to develop and validate fuzzy Bayes models to identify print defects, an experiment was conducted in Korea for three weeks from November 21 to December 8, 2006. The experiment was intended to collect English and Korean description data by Korean subjects that describe 32 representative print defects of laser printers manufactured by one company. The Bayes models obtained from the experiment will be implemented in a query search function on Websites for self-help print quality troubleshooting.

2.1   Subjects

A total of 42 Korean subjects applied for the experiment. However, only 40 subjects participated in the experiment because two subjects were found to be colorblind from the Ishihara color test [4]. The data from the first 30 subjects were used to develop fuzzy Bayes models, and the data from the last ten subjects were used to evaluate the fuzzy Bayes models. The subjects were composed of 37 males and 3 females, all of which are undergraduate, graduate students, or a post doctorial researcher in engineering or business schools from five universities in Korea (Hanyang University

at Seoul, Hanyang Univeristy at Ansan, POSTECH, Pukyong National University, Kumoh Natioal Institute of Technology). All subjects had TOEIC (Test of English for International Communication) scores greater than or equal to 550 (Average: 702), which was intended to qualify subjects who were assumed to read information written in English on the Web sites.

2.2 Procedure

The experiments were conducted under fluorescent illumination in five different laboratories located at five universities, which are similar to a normal office using laser printers. As a formal procedure, subjects signed a consent form approving that they would participate the experiment. Prior to the experiment, a colorblindness test

Table 1. Thirty Two Print Defects

Defect	Defect description	Background Color	Orientation	Pattern
D_1	Horizontal lines repeating every 1.4 inches	C	H	P_1
D_2	Horizontal lines repeating every 1.4 inches	M	H	P_1
D_3	Horizontal lines repeating every 1.4 inches	Y	H	P_1
D_4	Horizontal lines repeating every 1.4 inches	K	H	P_1
D_5	Horizontal lines repeating every 3.7 inches	C	H	P_2
D_6	Horizontal lines repeating every 3.7 inches	M	H	P_2
D_7	Horizontal lines repeating every 3.7 inches	Y	H	P_2
D_8	Horizontal lines repeating every 3.7 inches	K	H	P_2
D_9	Broad fuzzy horizontal bands repeating every 1.4 inches	C	H	P_3
D_{10}	Broad fuzzy horizontal bands repeating every 1.4 inches	M	H	P_3
D_{11}	Broad fuzzy horizontal bands repeating every 1.4 inches	Y	H	P_3
D_{12}	Broad fuzzy horizontal bands repeating every 1.4 inches	K	H	P_3
D_{13}	White vertical lines or bands	C	V	P_4
D_{14}	White vertical lines or bands	M	V	P_4
D_{15}	White vertical lines or bands	Y	V	P_4
D_{16}	White vertical lines or bands	K	V	P_4
D_{17}	Vertical lines or line segments	C	V	P_5
D_{18}	Vertical lines or line segments	M	V	P_5
D_{19}	Vertical lines or line segments	Y	V	P_5
D_{20}	Vertical lines or line segments	K	V	P_5
D_{21}	Fine vertical lines	C	V	P_6
D_{22}	Fine vertical lines	M	V	P_6
D_{23}	Fine vertical lines	Y	V	P_6
D_{24}	Fine vertical lines	K	V	P_6
D_{25}	Missing color	C	X	P_7
D_{26}	Missing color	M	X	P_7
D_{27}	Missing color	Y	X	P_7
D_{28}	Missing color	K	X	P_7
D_{29}	Tinted background	C	X	P_8
D_{30}	Tinted background	M	X	P_8
D_{31}	Tinted background	Y	X	P_8
D_{32}	Tinted background	K	X	P_8

C: cyan, M: magenta, Y: yellow, K: black; H: horizontal, V: vertical, X: others; P_i: pattern i

was conducted because print defects in the experiment were closely related to colors. The Ishihara color test, which consists of a series of pictures of colored spots, was used to screen the colorblind subjects. A pretest questionnaire which was designed to measure the subjects' general knowledge about printers, print defects, and troubleshooting on the Web sites, was provided to the subjects at the beginning of the experiment. During the experiment, a pair of printouts and an answer sheet were provided to the subjects in random order to eliminate learning effects. One of each pair had no print defect and the other had a defect so that the subjects could compare the two pages and describe the defect. As shown in Table 1 (Refer to www.hp.com/cpso-support-new/pq/4600/home.html), there were a total of 32 print defects which had 8 different print defect patterns for four different colors, cyan (C), magenta (M), yellow (Y), and black (K). Among these print defects, "Horizontal lines repeating every 1.4 inches (P_1)", "Horizontal lines repeating every 3.7 inches (P_2)", and "Broad fuzzy horizontal bands repeating every 1.4 inches (P_3)" for any color are classified into horizontal print defects (H). "White vertical lines or bands (P_4)", "Vertical lines or line segments (P_5)", and "Fine vertical lines (P_6)" for any color are classified into vertical print defects (V). "Missing color (P_7)" and "Tinted background (P_8)" for any color are classified into others (X) on the print quality troubleshooting Web site. After comparing a pair of printouts, the subjects described on the answer sheet how they perceived the print defect in Korean or English. After describing all defects in one language, they followed the same process with another set of printouts in the other language in random orders from subject to subject. From the experiment, 1280 descriptions in English and 1280 descriptions in Korean were obtained. Out of each set of 1280 data, 960 were used for training and 320 were used for predictions.

3   Model Development

3.1   Parsing

Parsing is the process of using grammar to assign a syntactic analysis to a string of words, that is, the process of recognizing a sentence and simultaneously building a representation of its structure [11], [13]. The most common and easiest methods to implement parsing are based on keywords by chunking that are used to build sets of attribute values [9], [13]. For this study, shallow parsing analysis [9], [13] was conducted using the Textminer [7] which Purdue University has developed. Every word in the descriptions was separated by a space as shown in Figure 1.

ID	word	Tpred
15	yellow	t
15	line	t
15	empty	t
16	black	t

	ID	word	Tpred
▶	92	회색	t
	92	바탕	t
	93	파란	t
	93	프린트	t

(a) English word list example   (b) Korean word list example

Fig. 1. Word list examples

From the word lists, some non-essential words not helpful to identify print defects such as "is", "are", "and", and "of" in English, and "거", "겄", and "한다" in Korean were removed for a convenient analysis.

3.2 Lexical Analysis

Lexical analysis is used to divide a sentence into keywords and recognize the same word in its various possible forms [9]. In this study, only morphological similarity was considered and semantic similarity was not considered. For example, "line and lines", band and bands", "dark and darkness", or "add and added" are categorized into the same morphs. On the contrary, "error and fail" in English, or "붉은 and 빨간" in Korean are semantically similar, but morphologically different. Therefore, these words are categorized into different morphs. A total of 203 morphs in English and 304 morphs in Korean were created from the analysis. The result shows that Korean subjects used more morphs in Korean than those in English to describe all the print defects.

3.3 Fuzzy Bayes Model

The fuzzy Bayes model has been studied as a very useful technique to index new annotations to be added to an annotated bibliography [8] or to assign index terms or links to documents in a hypertext system [14]. The relation between descriptions and possible categories of print defects can be quantified using the expression.

$$P(D_i \mid w) = MAX_j \, P(D_i \mid w_j) = MAX_j \, \frac{P(w_j \mid D_i)P(D_i)}{P(w_j)}, \ (i=1,2,\ldots, 32 \ \text{and} \ j=1,2,\ldots,q) \ (1)$$

where q is the number of words in the description. $P(D_i/w)$ refers to the probability of a category of print defect I_i given the word present in the description. $P(w_j/D_i)$ is the probability of word w_j given a category of print defect D_i. $P(D_i)$ is the probability of a category of print defect I_i. $P(w_j)$ is probability of a particular word w_j in the description. The operator Max_j then selects the maximum probability calculated over the j words. The model calculated the probability of all keywords being included in certain categories of print defects as shown in Figure 2. The keywords include single words (Figure 2a) and word combinations (Figure 2b) in the word list. Paired words, triplet words, and quadruplet words were considered for the word combinations because the previous study [9], [10] insisted that word combinations greater than word quadruplets were not helpful in identifying print defect categories. In addition, the keywords for the analysis were included in case that the frequency of single words or any word in the word combinations is greater than four. With those rules, description data for both Korean and English training set was reduced to 955. Analysis of the descriptions showed that certain keywords were strongly associated with certain categories of print defects. The strength of these associations was described using mean and Likelihood Ratios as shown in the Figure 2. The likelihood Ratio (LR) for keywords was calculated by taking the ratio of the probability of word w_j given the print defect category D_i divided by the probability of word w_j. That is, $LR(D_i/ w_j) = P(w_j/D_i)/P(w_j)$, and $P(D_i/ w_j) = LR(D_i/ w_j) \times P(D_i)$ from equation (1).

Consequently, the LR is proportional to the Bayes estimator which also can be calculated by taking the ratio of frequency within a defect category divided by total frequency within the whole data set of a keyword. Taking Figure 2a as an example, the word "black" appeared 53 times out of the total 955 descriptions, and 7 times out of 30 of D_{32}. Thus, the likelihood ratio (LR) for the "black" is calculated as follows:

LR(D_{32}/black) = P(black/ D_{32})/P(black) = (7/30)/(53/955) = 4.20
P(D_{32}/black) = LR(D_{32}/black) x P (D_{32}) = 4.20 x (30/955) = 7/53 = 0.13.

The higher the LR is, the higher the accuracy in identifying print defects.

word	defect	frequency	Total	mean	LR
a	8	3	14	0,21428571429	7,05665024630542
all	28	2	5	0,4	12,7333333333333
black	32	7	53	0,1320754717	4,20440251572327
black	4	9	53	0,16981132075	5,59206245933637

(a) Example for English single word

word	defect	frequency	Total	mean	LR
빨간&가로&개	6	2	4	0,5	15,9666666666667
빨간&가로&개	2	2	4	0,5	15,9666666666667
빨간&가로&선	6	2	4	0,5	15,9666666666667
빨간&가로&줄	2	7	15	0,46666667	14,9022222222222

(b) Example for Korean word combination

Fig. 2. Examples of keywords and Bayes estimators

4   Model Validation

4.1   Model Performance Within Top Five Predictions

As mentioned before, 320 additional descriptions were used to evaluate a fuzzy Bayes model developed by 955 descriptions. Applying each description to the Bayes model, LR was calculated for all keywords of English and Korean and sorted according to the LR. Table 2 shows an example of LR listed in descending order for one description, "added some lines of different color" which was a description of D_{23} (Yellow fine vertical lines). The first keyword of the description is a paired word, "ink & yellow", and it predicts D_{31} (Yellow tinted background) with LR = 12.73, which is wrong. The second keyword of the description is a word triplet, "line & vertical & yellow", and it predicts D_{23} (Yellow fine vertical lines) with LR = 11.50, which is correct.

By the same process, 320 descriptions for prediction were analyzed in the aspect of perdition accuracy. Table 3 shows that the English model successfully predicted 26%, 46%, 54%, 60%, and 66%, and Korean model successfully predicted 33%, 54%, 66%, 72%, and 78% within the first, the second, the third, the fourth, and the fifth

predictions. Overall percentages of correct prediction rates were much decreased compared with those of previous study [9], [10], 50%, 66, 70%, 77%, and 81%. The main two reasons for the difference are as follows: (1) The number of print defects was increased from 16 in the previous study to 32 of this study; (2) there were many similarities among the 32 defects in this study. For example, D_1, D_2, D_3, and D_4 are exactly the same defects except for their colors. Print defects, D_1, and D_5 are very similar except for the distance between lines. However, subjects often failed to notice some critical factors such as the color and the distance.

Another result is that there was significant difference between English and Korean predictions, which showed that there were some problems when Korean subjects used English queries to find information and which supports the need for research into a Q-KE-CLD (query translation based Korean-English cross language diagnosis) system. The main reason for the difference between English and Korean predictions by Korean subjects was that when they described print defects in English, they tried to make the descriptions simple due to their lack of ability to describe them in English.

Table 2. Fuzzy Bayes Estimation Based on LR

Keyword	Prediction	Frequency	Total	Mean	LR
ink&yellow	31	2	5	0.16	12.73
line&vertical&yellow	23	13	36	0.30	11.50
vertical&yellow	23	14	41	0.28	10.87
line&vertical&yellow	15	12	36	0.30	10.61
vertical&yellow	15	13	41	0.28	10.09
line&vertical&yellow	19	9	36	0.30	7.96
vertical&yellow	19	10	41	0.28	7.76
line&yellow	3	24	111	0.17	6.88
line&yellow	7	24	111	0.17	6.88
line&yellow	23	19	111	0.17	5.45
line&yellow	15	17	111	0.17	4.88
yellow	3	30	208	0.11	4.59

Table 3. Model Accuracies within Top Five Predictions

Prediction Language	Correct Prediction	Top	Top Two	Top Three	Top Four	Top Five
English Predictions	Number (out of 320)	84	146	172	193	210
	Ratio	26%	46%	54%	60%	66%
Korean Predictions	Number (out of 320)	104	173	211	229	249
	Ratio	33%	54%	66%	72%	78%
Previous Study (Leman, 2003)	Number (out of 64)	32	42	45	49	52
	Ratio	50%	66%	70%	77%	81%

As mentioned previously, 203 morphs for English descriptions and 304 morphs for Korean descriptions were used. That is to say, when they described print defects in English, they missed many critical features in the print defects such as color, defect orientation, thickness, and distance. Therefore, in the case of English predictions, there were relatively fewer keywords to identify the print defects.

4.2   Model Performance By Print Defect Features

As mentioned before, colors, orientation, and patterns of the defects are the main features used to identify print defects. In this respect, the data was also re-classified based on the main features for both English and Korean descriptions. Figure 3 shows the prediction of defect orientation in English using the Textminer [7] as an example of predicting the main defect features.

ID	category	prediction	accuracy	predictor
1004	H	H	1	horizon
1005	V	X	0	dot&white
1006	V	V	1	stripe&white
1007	V	V	1	two&white

Fig. 3. Predicting defect orientations by English

Table 4 summarizes the accuracies of the first top predictions when the fuzzy Bayes model predicted colors, orientations, and patterns of the print defects with both Korean and English descriptions. As shown in Table 4, in general, the prediction accuracy of each of the main features was much higher than the accuracy of print defects of English (26%) and Korean (33%). Moreover, any accuracy in predicting the main features was higher than the accuracy (50%) in the previous study [10]. Especially, the prediction accuracy for defect orientations is over 90% for both English (90%) and Korean (94%). In the case of defect colors, Korean predictions (87%) were much more accurate than English prediction (55%), which means that Korean subjects were much worse at describing features of defect colors when they used English. In the case of defect patterns, English predictions (61%) were a bit higher than Korean predictions (57%), but rather close. As expected, multiplying all

Table 4. Model Accuracies by Main Features of Defects

Defect features	English prediction	Korean prediction
Colors (C, M, Y, K)	55% (177 out of 320)	87% (278 out of 320)
Orientations (H, V, X)	90% (288 out of 320)	94% (302 out of 320)
Patterns (P_1, P_2,…, P_7)	61% (194 out of 320)	57% (181 out of 320)
Multiplying three accuracies	30% (0.55 x 0.9 x 0.61)	47% (0.87 x 0.94 x 0.57)
Defects (D_1, D_2,…, D_{32})	26% (84 out of 320)	33% (104 out of 320)

accuracies for the main features (30% for English and 47% for Korean) was much higher than the accuracy for print defects because there were many other factors influencing the prediction accuracy such as line thickness, distance between lines, and number of lines.

5   Conclusions

The fuzzy Bayes model correctly predicted 26%, 46%, 54%, 60%, and 66% for English descriptions and 33%, 54%, 66%, 72%, and 78% for Korean descriptions within the first, the second, the third, the fourth, and the fifth predictions by Korean subjects. The percentage of correct prediction of Korean descriptions was much higher than that of English because Korean subjects tried to make English descriptions simpler due to their reluctance to use English, though the subject selection was restricted to individuals who have TOEIC scores over 550. In addition, a result shows that accuracies of predicting colors, orientations, and patterns of print defects were 55%, 90%, 61% respectively by English predictions, and 87%, 94%, 67% by Korean predictions, which were much higher than the accuracies respectively (26% by English, 33% by Korean) of predicting print defects.

6   Further Study

As mentioned before, the objective of this study is to investigate a Q-KE-CLD system. Therefore English description data by US subjects will be collected and a fuzzy Bayes model from the data will be developed. English descriptions by Korean subjects will be evaluated based on the fuzzy Bayes model developed with English descriptions by US subjects. In addition, Korean descriptions will be translated into English and evaluated based on the fuzzy Bayes model developed with English descriptions by US subjects. Finally, the Q-KE-CLD system will be developed and evaluated based on the fuzzy Bayes models from data obtained by Korean and US subjects.

References

1. Chen, J.: The construction, use, and evaluation of a lexical knowledge base for English-Chinese cross-language information retrieval. Doctoral dissertation, Syracuse University (2003)
2. Choe, P., Kim, C., Lehto, M.R., Lehto, X., Allebach, J.: Evaluating and improving a self-help technical support Web site: Use of focus group interviews. International Journal of Human-Computer Interaction 21(3), 333–354 (2006)

3. Diekema, A.R.: Translation events in cross-language information retrieval: Lexical ambiguity, lexical holes, vocabulary mismatch, and correct translations. Doctoral dissertation, Syracuse University (2003)
4. Ishihara, S.: 1996. Ishihara's tests for colour-defficiency. 38 plates edition (2003)
5. Park, H.J., Santos, H., Kim, C., Choe, P., Kumontoy, R., Low, K., Oldenberger, K., Ortiz, M., Lehto, X., Lehto, M.: A Web-based self-diagnosis tool to solve print quality issues. Proceedings of NIP22 22nd International Conference on Digital Printing Technologies, Denver, CO, September 17-22, 2006, pp. 465–471 (2006)
6. Kim, C., Choe, P., Lehto, M.R., Allebach, J.: Development of a web-based interactive self-help troubleshooting tool for print quality problems. In: International Conference on Human-Computer Interaction, Las Vegas, Nevada USA, July 22-27, 2005, pp. 22–27 (2005)
7. Lehto, M.R.: TextMiner Manual (Ann Arbor, MI: Consumer Research,Inc.) (2004)
8. Lehto, M.R.: Warnings and safety instructions: electronic hypertext version 2.0. Fuller technical publications, Ann Arbor, MI (1994)
9. Leman, S.: Decision support system to predict print quality. Master thesis, Purdue University (2001)
10. Leman, S., Lehto, M.: Interactive decision support system to predict print quality. Ergonomics 46(1-3), 52–67 (2003)
11. Mitkov, R.: The Oxford handbook of computational linguistics. Oxford University Press, Oxford, New York (2003)
12. Salton, G.: Experiments in multi-lingual information retrieval. Information Processing Letters 2(1), 6–11 (1973)
13. Smith, G.W.: Computers and human language. Oxford University Press, Oxford, New York (1991)
14. Zhu, W.: Decision support for indexing and retrieval of information in hypertext systems. International Journal of Human-Computer Interaction 11(4), 349–371 (1999)