

デジタル 4R 訓練システム構築のための成否判定システムの最適化手法の研究 Research of a method to improve performance to determine whether answer is right or not to realize a digital 4 round training system

箕輪弘嗣[†] 竹内孔一[‡] 藤本宏美[§]
Hirotsugu Minowa[†] Koichi Takeuchi[‡] Hiromi Fujimoto[§]

1. はじめに

産業における安全の向上が求められている。そのため、近年は仮想現実感(Virtual Reality, 以下 VR)を利用した訓練が期待されている。仮想環境における訓練では、訓練対象の本質を再現しての模擬訓練が可能となるが、基礎が身についていなければ習得効果が上がらない可能性がある。そこで基礎を自学自習する手段として e-learning に注目している。e-learning といった ICT 利活用により、教科書で学ぶ以上の学習機会が得られる事が期待できる。ただし、現在の e-learning は訓練者の理解力を十分に測る事ができない問題がある。それは、e-learning は選択形式であるため、記憶が曖昧でも回答できてしまえる事である。

そこで、機械学習を用い訓練者の力をより正確に測る事ができるよう、自由記述で回答を求める事で訓練者の力をより正確に評価する事に主眼をおいたデジタル 4R システムの開発[1][2]を目指している。本論文では、開発中のシステム[1]の中で用いる訓練者の回答の正否を判定する正否判定サブシステムの判定精度向上のための最適化法の研究の進捗について報告する。

2. 研究内容

正否判定サブシステムの最適化のため、正否判定サブシステム生成の訓練データを後述する通り工夫した手順で生成し、その時の診断性能を評価した。

2.1 回答事例の取得

交通に関する KYT シート集[3][4] からそれぞれ 7 枚の KYT シートを選び、1 グループ 2 人または 3 人に分かれた高校生 21 グループ (計 43 名) に回答してもらった。KYT とは危険予知訓練(Kiken Yochi Training)の略記である。この時、1 シートに対して 3 グループを割り当て、各グループは約 10 個ずつ、KYT シートのイラストに対する回答文を作成してもらった。この 7 枚のシートの選定理由は、本 KYT シートの中でも平易な内容で、運転免許を持たない高校生でも回答は困難でないだろうと判断したためである。高校生が抽出した潜在的危険を説明した回答文と筆者らが考えた模範解答を含む 30-40 個の文章を、筆者らが正例、負例に振り分けた。このデータから 2.2 節で述べる手順により訓練データを作成する。この訓練データにおける正例、負例のレコード数はほぼ同じであった。

2.2 訓練データ生成手順

Step 1. 名詞の抽出

形態素解析を実施して、名詞と解析された単語を抽出し、訓練データの素性とする。

Step 2. Step 1 の単語を含む WordNet ラベル抽出

WordNet から Step 1 で抽出した名詞と合致する単語を抽出し、訓練データの素性とする。抽出した WordNet のラベル全てを素性とした。

Step 3. 主語・述語の抽出

構文解析を利用して回答文から主語と述語を抽出し、訓練データの素性とする。その表記は“動詞原形_左記動詞に係っている名詞”とした。主語かどうかは、動詞に係り、「に/を/が」を助詞にもつ名詞であるか否かを条件とした。

Step 4. Step 3 の単語に合った WordNet ラベル抽出

WordNet から Step 3 で抽出した主語、述語の原形と合致する単語を抽出し、訓練データの素性とする。Step 3 の“動詞原形”と“左記動詞に係っている名詞”に該当する WordNet 上の 2 種類のラベルの直積の組み合わせを素性とする。

3. 結果

2.2 節に従い生成した訓練データを元に診断器を生成し、3.1 に示す評価方法で評価した。

3.1 評価方法

性能評価は 10 Cross Validation で実施した。レコードを約 3~4 個を含む 10 組みに分け SVM と Random forest で評価した。その判定スコアを 3.2 に示す。

3.2 判定スコアの評価結果

3.1 に従い SVM で解析した結果は表 1 の通りになった。

表 1 解析結果

#	シート名	シート NO.	比較元正答率 [%]※1[1]	提案手法の正答率 [%]	※2
1	交通編[3]	1	55	68.33	○
2		4	60.83	70	○
3		5	59.17	75.83	○
4		10	80	80.83	○
5		12	81.67	84.17	○
6		13	64.17	62.5	×
7		14	42.5	38.33	×
8	すぐに使える KYT シート集 [4]	3	69.17	71.67	○
9		11	90	86.67	×
10		16	82.5	82.5	×
11		17	52.5	55	○
12		18	53.33	57.5	○
13		19	80	89.17	○
14		20	72.5	68.33	×

※1: 訓練データの素性を回答文の名詞だけとした

※2: 提案手法で正答率が向上したケースは○

3.1 に従い Random Forest で解析した結果は表 2 の通りになった。

表 2 解析結果

#	シート名	シート NO.	比較元正答率 [%]※1	提案手法の正答率 [%]※2	
1	交通編[3]	1	74.17	70	×
2		4	58.33	66.67	○
3		5	61.68	65.01	○
4		10	67.5	75.01	○
5		12	67.51	76.67	○
6		13	68.34	59.17	×
7		14	39.16	51.67	○
8	すぐに使える KYT シート集 [4]	3	69.17	66.67	×
9		11	90.01	86.68	×
10		16	74.17	80	○
11		17	62.5	60	×
12		18	71.67	70	×
13		19	72.51	86.67	○
14		20	48.34	52.5	○

※1: 訓練データの素性を回答文の名詞だけとした

※2: 提案手法で正答率向上したケースは○

3.3 主語・述語の抽出率

Step 3 において MeCab[5], Cabocha[6]を使って抽出した主語・述語の正答率を表 3, 表 4 に示す。

表 3 主語・述語の抽出率-交通編

#	シート名	シート NO.	主語 / 述語	正解数	不正解数	正答率 [%]	平均正答率
1		1	主語	26	5	84	89
			述語	29	2	94	
2		4	主語	23	14	62	69
			述語	28	9	76	
3		5	主語	25	9	74	76
			述語	27	7	79	
4	交通編 [3]	10	主語	21	13	62	69
			述語	26	8	76	
5		12	主語	25	12	68	81
			述語	35	2	95	
6		13	主語	23	8	74	82
			述語	28	3	90	
7		14	主語	27	10	73	78
			述語	31	6	84	

表 4 主語・述語の抽出率-すぐに使える KYT シート集

#	シート名	シート NO.	主語 / 述語	正解数	不正解数	正答率 [%]	平均正答率
8		3	主語	33	6	85	82
			述語	31	8	79	
9	すぐに使える KYT シート集 [4]	11	主語	26	4	87	82
			述語	23	7	77	
10		16	主語	29	9	76	82
			述語	33	5	87	
11		17	主語	34	6	85	76
			述語	28	12	70	
12		18	主語	33	6	85	83

13	19	述語	32	7	82	80
		主語	23	12	66	
14	20	述語	33	2	94	76
		主語	29	8	78	
		述語	27	10	73	

3.4 統計的有意性評価

統計的有意差を測るため、Wilcoxon 符号順位検定と T 検定で評価した。解析結果は定量的であるため、T 検定を使う事が望ましいが、下記に述べる通り前提条件がある。T 検定の前提条件には、①標本は母集団から無作為に抽出されていること。②母集団の分布が正規分布であること。③2つの母集団の分散が等しい(等分散性)、が挙げられる。

前提条件に従っていない可能性を懸念し、Wilcoxon 符号順位検定でも評価した。Wilcoxon 符号順位検定は、1対の標本に対するノンパラメトリック検定法である。

Wilcoxon 符号順位検定において P 値は 0.1306>5%であり、優位性は認められなかった。T 検定と違い、両者の差による符号だけで評価した結果のためと考えられる。

T 検定では、P 値は 0.1796>5%となり、こちらでも優位性は認められなかった。

4. 考察

主語述語の抽出率とサブ診断器の正答率の相関は低そうである。正答率が低い原因と思われるのが、危険要素が複数含むシートでは回答の方向性が分散してしまう事である。回答数を増やし、意図しない回答を切り捨ててしまうなど対応が必要であろう。

5. おわりに

自由記述による e-learning の実現に向けて、危険予知訓練(KYT)シートにおける回答といった短文を SVM, Random Forest などの機械学習で高精度に正否判定できるように研究した結果を報告した。

今後、回答の分散性の評価方法について検討し、判定精度との相関を調べるなど、判定精度の改善法の研究に努める。

参考文献

- [1] Hirosugu Minowa, Hiromi Fujimoto, Koichi Takeuchi, "A Mechanical Method for Evaluating Trainee Answers in a Risk Prediction Training System Based on the 4R Method", *IEE*, vol. 1, no. 3, pp. 59-68 (2015).
- [2] H. Minowa, H. Fujimoto, K. Takeuchi, "Automatic Evaluation Methods of Trainee's Answers to Develop a 4R Risk Prediction Training System", *Proc. of the 2015 ILAI 4th International Congress on Advanced Applied Informatics*, pp. 283-286 (2015).
- [3] 中央労働災害防止協会, "危険予知訓練(KYT) イラストシート集-交通編", 中央労働災害防止協会 (2013).
- [4] 中央労働災害防止協会, "すぐに使える KYT イラストシート集 1 - みんなでやろう! 危険予知訓練-", 中央労働災害防止協会 (2014).
- [5] T. Kudo, K. Yamamoto, Y. Matsumoto, "Applying Conditional Random Fields to Japanese Morphological Analysis", *Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp. 230-237, (2004).
- [6] T. Kudo, Y. Matsumoto, "Japanese Dependency Analysis using Cascaded Chunking", *CoNLL 2002: Proc. of the 6th Conference on Natural Language Learning 2002*, pp. 63-69 (2002).