

交通規則問題のための解答システムの構築

Constructing Solvers for Traffic Rule Questions

杉村 皓太 佐々木 裕
Kota Sugimura Yutaka Sasaki

豊田工業大学
Toyota Technological Institute

1 はじめに

質問応答システムに関する一連の研究のなかで、クイズに答えるシステムの研究や大学入試問題を解くシステムの研究が行われてきた。2011年には、IBMのワトソンシステムが米国のクイズ番組において全米チャンピオン2名を破ったことが、大きく報道された。

しかしながら、これまで自動車免許を取得するために課せられる交通規則に関する学科試験の問題を解くシステムの研究はほとんど行われていない。中川ら[1]は、1990年代に学科試験問題文を論理式に変換する先駆的な研究を行ったが、幅広い交通問題を解くには至らなかった。本研究では、近年の質問応答研究のアプローチに沿って、用語の共起や一致に関する計量情報に基づいて問題を解く手法について報告する。

2 学科試験問題

自動車の運転免許を取得する際には、学科試験を受験し合格する必要がある。学科試験の問題は、交通規則に関する95問の正誤問題であり90点以上で合格となる。ただし、本研究では、イラストや道路標識といった図の認識は対象としない。

問題の例を以下に示す。

Q 普通自動車免許で小型特殊自動車を運転することができる。

A ○

問題の内容は、狭義の交通規則に関する知識を問うものに限らず、下記のような常識問題も含まれる。

Q 右カーブでは、右方向に遠心力が働く。

A ×

また、問題文の否定表現を扱う必要がある。

Q 普通自動車免許では原動機付自転車は運転できない。

A ×

本研究は、究極的には学科試験の合格ラインを超えることを目指しているが、本報告では、プロトタイプシステムの現状について報告する。

3 研究の流れ

本研究は、以下の4つの構成要素からなる。

- (1) 交通に関する用語の収集
- (2) 交通用語ソーラスの構築
- (3) 交通に関する問題の収集
- (4) 問題を解くシステムの作成

3.1 交通に関する用語の収集

交通に関する用語を文献[2-5]から人手により収集した。現段階では、約1500語の用語が収集されている。収集した単語数を表1に示す。

表1: 収集した単語数

	単語数
動詞	13
数値	69
名詞	1447
合計	1529

3.2 交通シソーラスの構築

収集した用語を人手により体系化し，交通シソーラスを作成した。

シソーラスの上位は以下のような形をしている。

交通知識

- 抽象物
 - 色
 - 単位
 - 現象
 - 罰則
 - …
- 具体物
 - 移動体
 - ◇ 自動車
 - ◇ 軽車両
 - ◇ 電車
 - ◇ バス
 - ◇ …
 - 固定物
 - ◇ 道路
 - ◇ 信号機
 - ◇ 標識
 - ◇ …
- 動作
 - 割り込む
 - 撥ねる
 - …
- 人
 - 大人
 - 子供
 - 歩行者
 - …

収集した交通用語には，ID を割り当てた。交通用語はシソーラスのノードに割り当てられており，用語 ID はシソーラスの体系のルートから対象ノードに至るパスをあらわしたものに，用語自体の ID を加えたものである。たとえば，用語 W がノード root→A→B の下に割り当てられている場合，ID は root. A. B. n の形式で与えられる。

用語は，ID とともに MeCab のユーザ辞書に登録し，MeCab や Cabocha による形態素解析

時に交通用語であることを判定できるようにした。

3.3 交通に関する問題の収集

交通法規問題を開発用に約 700 問，評価用に約 150 問[6-9]から収集した。利用する問題は日本語で表されている文章問題のみで，標識や状況判断など，図を用いた問題は考慮していない。また，練習用の問題，評価用の問題どちらも解答が「正」の問題と「誤」の問題の割合はおよそ 5 割である。問題の重複がないことを人手により確認した。収集した問題数を表 2 に示す。

表 2：収集した問題数

	問題数	
	開発用	評価用
正	333	70
誤	352	75
計	685	145

収集した問題はプログラム中で利用しやすいような XML 形式のコーパスとした。問題にはそれぞれ問題番号が付与されている。

```
<DOCS>
<DOC>
<DOCID>00001</DOCID>
<TEXT>
<Q>直面の信号が赤色の灯火で、同時に青色の矢印信号が左に出たときは、自動車や原動機付自転車は矢印の方向へ進行できるが、軽車両は進行してはならない。</Q>
<ANS>×</ANS>
</TEXT>
</DOC>
<DOC>
<DOCID>00002</DOCID>
<TEXT>
<Q>高速自動車国道の本線車道での大型貨物自動車の最高速度は100キロメートル毎時である。</Q>
<ANS>×</ANS>
</TEXT>
</DOC>
</DOCS>
```

<DOCID>タグで囲まれている部分は問題番号，<Q>タグで囲まれている部分は問題文，<ANS>

と</ANS>で囲まれている部分はその問題の正解を表している。

3.4 交通法規問題を解くシステム

作成したシステムのブロック図を図1に示す。

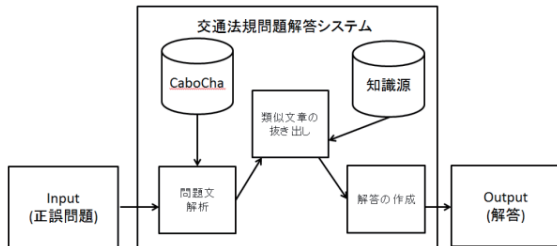


図1: 解答システムブロック図

解答作成の基本的な流れは次の通りである。

- (1) 問題文章を構文解析し、形態素及び文節の形に分解する。形態素は任意の品詞の形態素を対象形態素として処理に用いる。
- (2) 分解した各対象形態素及び文節を含む文章を交通教則コーパスから検索し、文章がピックアップされた回数をカウントする。
- (3) 問題文章中の知識源に含まれている交通用語と、長さLの単語を抽出する。
- (4) 抽出単語が含まれており、(2)のカウント数が一番多い文章を類似候補文とする。
- (5) 類似候補文について、問題文章の対象形態素がどれだけ含まれているかを算出

$$\text{類似度 } \text{Sim} = \frac{n_i}{N_i}$$

n_i : 類似候補文に含まれる対象形態素数

N_i : 問題文の対象形態素数

- (6) Sim が任意の閾値以上のとき正(○)とし、それ以外を誤(×)とする。
- (7) 解答時に注意すべき単語群及び文体の肯定・否定情報から(6)の解を修正する。

図2に解答システム出力結果の一部を示す。

○ 路端から発進するときは、方向指示器などによって発進の合図をし、もう一度バックミラーなどで前後左右の安全を確かめなければならない。EOS候補			
方向指示器などによって発進の合図をし、もう一度バックミラーなどで前後左右の安全を確かめてから発進するEOS			
10	14	0.714285714285714	最終結果 ○

図2: 解答システム出力結果(一部)

図2 上部が問題の正解と問題文章、下部が類似候補文及び n_i , N_i , Sim, 判定結果である。対象形態素の一致確認には unigram と bigram を用いている。また、unigram については形態素の表層形と原形を対象形態素として用いており、bigram は形態素の原形を対象形態素として用いている。

(7)については、「徐行」や「一時停止」などの解答時にキーとなる単語群を用意しておく、判定した解答が○だった場合、問題文中のキー単語が類似候補文に含まれていなければ×と判定するという手法と、問題文と類似候補文の肯定・否定情報を語尾から取得し、問題文と類似候補文の肯定・否定情報が逆だった場合(6)にて判定した解答を反転させるといった手法を取った。

4 実験

交通問題を解くための知識源として、「交通の方法に関する教則」(国家公安委員会告示)を用いる。交通教則は運転に関する指針をまとめた行政文書であるが、法律文調ではなく、学科試験問題と同様な平易な表現で交通規則を記述しており、知識源として適切である。

以下に例を示す。

第1章 歩行者と運転者に共通の心得

車は、私たちの生活から切り離せない身近な文明の利器になっている。しかし、その反面、使い方を誤ると悲惨な交通事故を起こす凶器になったり、騒音、振動などにより沿道住民に大きな被害を及ぼす原因になったりする。また、自分勝手な通行の仕方がもとで争いが生じ、人間関係を険悪化させる場面も日常よく見受けられる。

車社会においては、歩行者も運転者もそれぞれの責任を自覚して、周りの人に迷惑を掛けず、安全、快適に通行することができるような交通環境をつくりあげるよう努めなければならない。そのためには、あらかじめ、車と交通について正しい知識を持ち、正しい交通の方法を身に付けておくとともに、実際の交通の場においても、自分本位でなく相手に対する思いやりの気持ちを持って、判断し、行動することが必要となる。

本システムにおいて、対象形態素を名詞、動詞、形容詞とし、(3)における単語の長さLを4文字として開発用問題を用いて実験を行った。形態素の原形を用いた場合の unigram と bigram の閾値と正答率の関係を以下の図3に示す。また、図3の正答率が最大のときの条件の下で表層形の unigram の閾値を変化させた実行結果を以下の表3に示す。

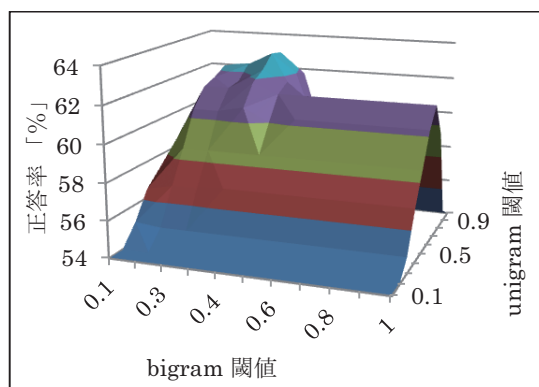


図3：原形を用いた場合の unigram と bigram の閾値と正答率の関係

表3：表層形を用いた場合の unigram 閾値の変化と正答率の関係

閾値変化	正答率 [%]
0.1	51.82
0.2	52.85
0.3	53.28
0.4	56.93
0.5	59.12
0.6	62.34
0.7	63.80
0.8	65.26
0.9	63.50
1.0	62.92

これらの結果から、(6)における原形の unigram の閾値を 0.9, bigram の閾値を 0.35, 表層形の unigram の閾値を 0.8 と決定した。

以上の条件より、開発用問題と評価用問題 についての実験結果を以下の表4に示す。

表4：解答システムの出力結果

問題	題数 [問]	正答率 [%]
評価用問題	145	66.21
開発用問題	685	65.26

5 まとめと今後の課題

システムの正答率としては、ベースラインとなる 50%に対して、16%程度の精度向上を実現した。しかしながら、合格ラインを達成するという目標には届いてはいない。

システムの出力結果を分析したところ、知識源の交通教則に問題文と類似した文章が見られる場合は、類似候補文は妥当に得られていると感じられた。今後正答率を向上させるためには、知識源をさらに拡張させる、係り受け解析の結果を用いる、機械学習を用いるなどが考えられる。

参考文献

- [1] 中川 聖一他：交通規則文に関する質問応答システム LICENCE における日本語文から一階述語論理式への変換，情報処理学会論文誌，Vol. 32, No. 3, pp. 354-363, 1991.
- [2] 警察庁交通局（監修），全日本交通安全協会（編集）：交通の教則[運転者用]，全日本交通安全協会，2011.
- [3] 警察庁交通局（監修），全日本交通安全協会（編集）：人にやさしい安全運転，全日本交通安全協会，2011.
- [4] 中部日本自動車学校，学科教本，中部日本自動車学校.
- [5] トヨタ名古屋教育センター，運転教本.
- [6] 頑張れ！普通免許学科試験 (<http://homepage3.nifty.com/menkyo/index.html>)
- [7] 免許なび (<http://menkyo.ne.jp/>)
- [8] 運転免許どっととる (<http://www5b.biglobe.ne.jp/~nobusann/>)
- [9] インターネット自動車教習所 (<http://menkyo119.com/>)