

ファジィ集合の概念を用いた 日本語質問応答システムの試作

馬野 元秀

(大阪大学・大型計算機センター)

1. はじめに

自然言語には、使用する人間の柔軟さを反映して、さまざまな形のあいまいさが含まれている。自然言語の研究が最近盛んに行なわれ、統語的、意味的な解釈に対するあいまいさ（多義性）についてはよく研究されているが、単語がもつ意味のあいまいさについてはほとんど研究されていない。あいまいな意味をもつ単語をあいまいな形のままで理解し、あいまいな質問に対して答えるということは、我々人間がごく普通に行なっていることである。機械とのスムーズなインタフェースなどを考えると、このようなあいまいさの表現と取り扱いは非常に重要であると思われる。

そこで、文献[1]の日本語質問応答システム TQAS(Tiny Question Answering System)をもとにして「大きい」、「小さい」、「中くらい」、「とても大きい」、「まあ大きい」などのあいまいな概念を含む単語を用いることのできる簡単な日本語質問応答システムを試作した。システム内部では、単語のもつあいまいな概念を可能性分布[2]で表現し、あいまいな概念を含む単語による質問に対しては、質問中のファジィ集合またはファジィ関係[3]とシステム内部の可能性分布との満足度を計算し、これに基づいて返答するようにした。

以下、2節ではファジィ集合と可能性分布についてまとめ、3節で試作した日本語質問応答システムについて詳しく述べる。

2. ファジィ集合と可能性分布

ファジィ集合も可能性分布も全体集合から単位区間[0,1]への写像として定義される。「大きい」や「重い」などのように、ある要素が集合に属するか属さないかの境界がはっきりとわからないときに、その要素が集合に属する度合い（グレード）を考える。グレードは、完全に属するを1とし、完全に属さないを0として、その間の値で表わす。そして、要素とグレードの集まりFをファジィ集合（fuzzy set）と呼び、

$$\begin{aligned} F &= \{ \mu_F(u_1) / u_1, \mu_F(u_2) / u_2, \dots, \mu_F(u_n) / u_n \} \\ &= \{ \mu_F(u) / u : u \in U \} \end{aligned}$$

と書き表わす。このとき、要素からグレードへの関数をメンバシップ関数（membership function）と呼ぶ。

大きさについて0から10の尺度を考えたときに、ファジィ集合「大きい」は、例えば、

$$\text{大きい} = \{1/10, 0.92/9, 0.68/8, 0.32/7, 0.08/6\}$$

となる。

さらに、2変数のメンバシップ関数を持つファジィ集合を考えることができ、これはファジィ関係と

呼ばれる。

可能性分布は文 (assertion) が持っている情報を表現するために考えられたものである[2]。普通の文 (あいまいではない文) ，

A1 : 箱box1の大きさは 5か 6である。

を考えよう。他の情報がまったくないときに、この文から得られる情報は、

- (i) 箱box1の大きさは 5である可能性があり、 6である可能性もある。
- (ii) 箱box1の大きさが 5又は 6以外である可能性はない。

である。「可能性がある」を 1で「可能性がない」を 0で表わせば、箱box1の大きさについて、「大きさ 5と 6で可能性が 1, その他の大きさが可能性が 0である」という可能性の分布が得られる。これを 可能性分布 (possibility distribution) と呼び、

$$\Pi_{\text{大きさ(box1)}} = \{1/5, 1/6\} p$$

と表記する。ただし、「大きさ(box1)」は箱box1の大きさの属性を表わす。また、右辺の最後のp でフアジィ集合ではなく可能性分布であることを示す。

次に、文A1の代わりに、

A2 : 箱box2は大きい。

を考えよう。この場合には、箱box2の大きさに関する可能性分布は 0と 1だけでは表現できない。そこで、可能性の取りうる値を単位区間[0,1] に拡張して考えると、文A2から

$$\Pi_{\text{大きさ(box2)}} = \{1/10, 0.92/9, 0.68/8, 0.32/7, 0.08/6\} p$$

のような可能性分布を得ることができる。

一般に、「Xは Pである」という形の文からは、

$$\begin{aligned} \Pi_{A(X)} &= \{ \pi_{A(X)}(u_1) / u_1, \pi_{A(X)}(u_2) / u_2, \dots, \pi_{A(X)}(u_n) / u_n \} p \\ &= \{ \pi_{A(X)}(u) / u : u \in U \} p \end{aligned}$$

という可能性分布が得られる。ここで、A(X)は対象X のいま注目している (すなわち、P に関する) 属性を表わす。可能性分布において、フアジィ集合のメンバシップ関数 μ に当たる関数 π を 可能性分布関数 (possibility distribution function) と呼ぶ。

可能性分布は「Xは Pである」という形の文に結び付いており、対象X のある属性A について Xはある値を持っているが、いま分かっている情報からでは可能性分布 $\Pi_{A(X)}=P$ 以上は分からないという状況を表わしている。一方、フアジィ集合は通常の集合の特徴関数で値域を {0, 1} から [0,1]に拡張したもので、特徴関数が 0と 1の真理値上への関数と考えるならば、メンバシップ関数は[0,1] の真理値上への多値論理関数と考えられる。

3. フアジィ集合の概念を用いた日本語質問応答システム

本システムは、文献[1] の Lisp による小型日本語質問応答システム TQAS を VAX-11/780 (OS は VAX/VMS) 上のLisp-F3 に移植したものをベースとして、これにあいまいな概念を含む単語を処理する

部分を追加して実現した。システム内部では、単語のもつあいまいな概念を可能性分布[2]で表現し、あいまいな概念を含む単語による質問に対しては、質問中のファジィ集合とシステム内部の可能性分布との満足度を計算し、これに基づいて返答するようにした。

以下、システムとの会話例を示しながら、システムの概略について述べていこう。現在のところ端末機の都合によりローマ字しか使えないので、本稿ではその次の行に漢字かな混じり文を補ってある。

(1) 平叙文による記述

まず、どのような入力が可能か、例を見てみよう。>が付いた行が人間の入力で、*の付いた行がシステムからの応答である。

> akai hakoga tsukueno ueni arimasu.
(赤い箱が机の上にあります)

* wakarimashita
(わかりました)

> sono hakowa totemo omoi.
(その箱はとても重い)

* wakarimashita
(わかりました)

> aoi hakoga aru.
(青い箱がある)

* wakarimashita
(わかりました)

> sono hakowa omosaga chuukurai desu.
(その箱は重さが中くらいです)

* wakarimashita
(わかりました)

> kiiroi hanawa maa ookii.
(黄色い花はまあ大きい)

* wakarimashita
(わかりました)

> shiroi hanaga tsukueno ueni arimasu.
(白い花が机の上にあります)

* wakarimashita
(わかりました)

> sono hanawa chiisai.
(その花は小さい)

* wakarimashita
(わかりました)

> kuroi hakowa omosaga 7kurai desu.
(黒い箱は重さが7くらいです)

* wakarimashita
(わかりました)

現在のところ、辞書に登録されている名詞は、

hako (箱), hana (花), mari (まり), tsukue (机), tsumiki (積木)

の5つだけである。

名詞に付けることのできる属性には、

色, 大きさ, 重さ, 長さ

がある。色の属性で使用できる形容詞は、

akai (赤い), aoi (青い), kiiroi (黄色い), kuroi (黒い), shiroi (白い)

の5つで、現在のところ、これらはいまいでないと考えている。大きさ, 重さ, 長さの属性で使える形容詞は表1の通りで、これらはすべてあいまいであるとし、0から10までの11段階の尺度上での可能性分布と考える。そして、簡単のために同じ行の単語(例えば, ookii (大きい), omoi (重い), nagai (長い) など)は同じ形の可能性分布であるとしている。

表1. あいまいさを含む形容詞。

<大きさ>	<重さ>	<長さ>
ookii (大きい)	omoi (重い)	nagai (長い)
chiisai (小さい)	karui (軽い)	mijikai (短かい)
chuukurai (中くらい)		
nukurai (nくらい : $1 < n < 9$)		

そして、各単語の可能性分布としては、例えば、大きさの属性の形容詞に対して、

$$\text{ookii} = \{1/10, 0.92/9, 0.68/8, 0.32/7, 0.08/6\} p$$

$$\text{chiisai} = \{1/0, 0.92/1, 0.68/2, 0.32/3, 0.08/4\} p$$

$$\text{nukurai} = \{0.22/n-2, 0.78/n-1, 1/n, 0.78/n+1, 0.22/n+2\} p$$

と定義してある。そして、chuukurai は 5kurai と同じであるとしている。

また、色以外の属性の形容詞に対して使える副詞として、

totemo (とても), taihen (たいへん), maa (まあ), sukosi (少し)

がある。totemo と taihen は共に修飾する形容詞の可能性の各々の値を2乗し、maa と sukosi は0.5乗している[4]。

上の会話の結果、中期記憶の内容は、

tukue ---ue--> shiroi chiisai hana, akai totemo omoi hako
(机 --- 上 ---> 白い小さい花, 赤いととても重い箱)

aoi chuukuraino omosano hako
(青い中くらいの重さの箱)

kiiroi maa ookii hana
(黄色いまあ大きい花)

kuroi 7kuraino omosano hako
(黒い 7くらいの重さの箱)

となる。

(2) あいまいさを含まない質問

これについては、TQASと同じなので実行例をあげるにとどめる。

- > akai hakoga tsukuenu ueni arimasuka?
(赤い箱が机の上にありますか)
- * hai arimasu
(はい, あります)
- > naniga tsukuenu ueni arimasuka?
(何が机の上にありますか)
- * shiroi chiisai hana , akai totemo omoi hako --eeto kore dake desu
(白い小さい花, 赤いととも重い箱 - ええっと, これだけです)
- > naniirono hakoga arimasuka?
(何色の箱がありますか)
- * kuroiro , aoiro , akairo desu
(黒色, 青色, 赤色です)

(3) あいまいさを含む質問

まず, 実行例をいくつか示そう。

- > kiirono hanawa chiisai desuka?
(黄色い花は小さいですか)
- * iie chiisakuwa arimasen
(いいえ, 小さくはありません)
- > akai kakowa taihen omoi desuka?
(赤い箱はたいへん重いですか)
- * hai totemo omoi desu
(はい, とても重いです)
- > akai hakowa omoi desuka?
(赤い箱は重いですか)
- * hai tabun omoi deshou
(はい, たぶん重いでしょう)
- > akai hakowa maa omoi desuka?
(赤い箱はまあ重いですか)
- * hai tabun maa omoi deshou
(はい, たぶんまあ重いでしょう)

このような応答をするためには, 質問中の対象X についてのファジィ集合に対する, 中期記憶のなかの対象X に関する可能性分布 Π の満足度を求めることになる。満足度S は

$$S = \{ \pi(u) / \mu(u) : u \in U \} p$$

という区間 $[0,1]$ 上の可能性分布で与える。なお, U は $\{0,1,2,\dots,10\}$ と考えればよい。そして, こ

の結果をしきい値 θ (普通は0.5)と比べて, $\mu(u) > \theta$ となる $\pi(u)$ の最大値を π_1 とし, $\mu(u) < \theta$ となる $\pi(u)$ の最大値を π_0 とすると, 次の5つに分類できる.

- (a) 真であるT : $\pi_1 > 0, \pi_0 = 0$
- (b) たぶん真であるT' : $\pi_1 > 0, \pi_0 > 0, \pi_1 > \pi_0$
- (c) 偽であるF : $\pi_0 > 0, \pi_1 = 0$
- (d) たぶん偽であるF' : $\pi_0 > 0, \pi_1 > 0, \pi_0 > \pi_1$
- (e) わからないU : $\pi_0 = \pi_1$

ただし, これでは $\pi(u)$ と $\mu(u)$ の形が同じときにT'に分類されてしまうので, これに対しては別に調べてTにする. そして, これら(a)-(e)に基づいて, システムは

- (a) 真であるT : hai ... desu (はい, ...です)
- (b) たぶん真であるT' : hai tabun ... deshou (はい, たぶん...でしょう)
- (c) 偽であるF : iie ...dewa arimasen (いいえ, ...ではありません)
- (d) たぶん偽であるF' : iie tabun ...dewa naideshou (いいえ, たぶん...ではないでしょう)
- (e) わからないU : dochiratomo iemasen (どちらとも, いえません)

のように回答する.

例として, 「赤い箱は重いですか」について計算してみよう. 「赤い箱はとても重い」ので, 赤い箱の重さについての可能性分布は, omoi (重い) の可能性を2乗した

$$\Pi = \{1/10, 0.85/9, 0.46/8, 0.1/7\} p$$

が中期記憶に蓄えられている. 一方, 質問中のファジィ集合はomoi (重い) で, このメンバシップ関数は可能性分布omoi (重い) の可能性分布関数と同じ形の関数を使っているので,

$$F = \{1/10, 0.92/9, 0.68/8, 0.32/7, 0.08/6\}$$

となり, 満足度Sを計算すると

$$S = \{\pi(10)/\mu(10), \pi(9)/\mu(9), \pi(8)/\mu(8), \pi(7)/\mu(7), \pi(6)/\mu(6), \pi(5)/\mu(5)\} p$$

$$= \{1/1, 0.85/0.92, 0.46/0.68, 0.1/0.32, 0.08/0\} p$$

となる. そして, θ を0.5としているので, π_1 と π_0 は, それぞれ, 1と0.1となり, 「たぶん真であるT'」に分類される. したがって, システムは「hai tabun omoi deshou (はい, たぶん重いでしょう)」と回答する.

(4) 2つの対象についての質問

2つの対象を比べる質問も可能である. これは, 例えば,

- > akai hakowa aoi hakoyori zutto omoi desuka?
(赤い箱は青い箱よりずっと重いですか)
- * hai tabun zutto omoi deshou
(はい, たぶんずっと重いでしょう)

となる. また, hobo hitoshii (ほぼ等しい) という形容詞を使うこともできる.

> akai hakowa kuroi hakoto omosaga hobo hitoshii desuka?
 (赤い箱は黒い箱と重さがほぼ等しいですか)

* iie tabun hobo hitoshikuwa naideshou
 (いいえ、たぶんほぼ等しくはないでしょう)

> aoi hakowa kuroi hakoto omosaga habo hitoshii desuka?
 (青い箱は黒い箱と重さがほぼ等しいですか)

* hai tabun hobo hitoshii deshou
 (はい、たぶんほぼ等しいでしょう)

ここで、ファジィ関係 *zutto omoi* (ずっと重い) のメンバシップ関数は、

$$\mu_{>>}(u,v) = \begin{cases} 1 & u - v > 3 \\ 0.89 & u - v = 3 \\ 0.5 & u - v = 2 \\ 0.13 & u - v = 1 \\ 0 & \text{その他} \end{cases}$$

と定義してあり、ファジィ関係 *hobo hitoshii* (ほぼ等しい) のメンバシップ関数は、

$$\mu_{\sim}(u,v) = \begin{cases} 1 & u = v \\ 0.89 & |u - v| = 1 \\ 0.5 & |u - v| = 2 \\ 0.13 & |u - v| = 3 \\ 0 & |u - v| > 3 \end{cases}$$

と定義してある。

これは、質問のなかのファジィ関係 *R* に対する中期記憶のなかの可能性分布 Π_1 と Π_2 の満足度を求めることになる。この満足度 *S* は

$$S = \{ \min(\pi_1(u), \pi_2(v)) / \mu(u_1, u_2) : u \in U, v \in V \} p$$

という可能性分布で与えられる。これも区間 $[0,1]$ 上の可能性分布となるので、上と同じように、(a)-(e) に分類して回答をつくれればよい。

例として、「赤い箱は黒い箱と重さがほぼ等しいですか」の場合には、「赤い箱はとても重い」ので

$$\Pi_1 = \{ 1/10, 0.85/9, 0.46/8, 0.1/7 \} p$$

となり、「黒い箱は重さが7くらいです」だから、

$$\Pi_2 = \{ 0.22/5, 0.78/6, 1/7, 0.78/8, 0.22/9 \} p$$

となる。そして、質問中のファジィ集合は上の「ほぼ等しい」であるので、満足度は

$$S = \{ 1^{\wedge}0.22/0, 1^{\wedge}0.78/0, 1^{\wedge}1/0.13, 1^{\wedge}0.78/0.5, 1^{\wedge}0.22/0.89, 0.85^{\wedge}0.22/0, 0.85^{\wedge}0.78/0.13, 0.85^{\wedge}1/0.5, 0.85^{\wedge}0.78/0.89, 0.85^{\wedge}0.22/1, 0.46^{\wedge}0.22/0.13, 0.46^{\wedge}0.78/0.5, 0.46^{\wedge}1/0.89, 0.46^{\wedge}0.78/1, 0.46^{\wedge}0.22/0.89, 0.1^{\wedge}0.22/0.5, 0.1^{\wedge}0.78/0.89, 0.1^{\wedge}1/1, 0.1^{\wedge}0.78/0.89, 0.1^{\wedge}0.22/0.5 \} p$$

$$= \{0.78/0, 1/0.13, 0.85/0.5, 0.78/0.89, 0.46/1\} p$$

となる。ここで、 \wedge は minを表わす。そして θ が 0.5で、 π_1 と π_0 は、それぞれ、0.78と 1となり、「たぶん偽であるF'」に分類され、回答は「iee tabun hobo hitoshikuwa naideshou (いいえ、たぶんほぼ等しくはないでしょう)」となる。

3. おわりに

以上、VAX-11/780 上のLisp-F3 で作成したあいまいな概念を取り扱うことのできる日本語質問応答システムについて、例をあげて述べた。現在のところ、利用できる単語も多くないし、文法上の制約もかなりきびしいものになっている。今後は、より単語を増加させ、文法上の制約をゆるめて、より広い範囲の日本語を取り扱えるように拡張していきたい。

最後に、本システムは、筆者が岡山理科大学・理学部・応用数学科に勤務していたときに、当学科のVAX-11/780上に作成したものである。インプリメントについては、当時、当学科の学生であった吉沢延幸君と西野文子さんによるところが大である。深く感謝する。

[参考文献]

1. 田中, 元吉, 山梨: 「LISPで学ぶ認知心理学3 言語理解」, 東京大学出版会(1983).
2. L.A. Zadeh: "Fuzzy Sets as a Basis for a Theory of Possibility", Fuzzy Sets and Systems, Vol.1, pp.3-28 (1978).
3. L.A. Zadeh: "Fuzzy Sets", Information and Control, Vol.8, pp.338-353 (1965).
4. L.A. Zadeh: "A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges", Jour. of Cybernetics, Vol.2, pp.4-34 (1972).