

くらしの中の「情報」のサイエンス

～スマート社会を生きるための「情報」のすすめ～

ソフトウェア情報学部 学部長(教授) 猪股 俊光

1 はじめに

我が国が目指すべき未来社会の姿として政府により「Society 5.0」が提唱されています[1]。狩猟社会を「Society 1.0」、農耕社会を「Society 2.0」、工業社会を「Society 3.0」、情報社会を「Society 4.0」と位置づけ、「Society 5.0」は「超スマート社会」と位置づけています。

Society 5.0 が実現されると、どんな社会になるのでしょうか。「大量のデータから新たな価値が生まれる社会」へ変わります。そのためには『データ・数理科学』の研究領域が必要です。また、必要な情報が必要なときに提供される社会の実現のためには『人工知能』の研究分野が必要です。さらに、地域の課題に対応するために『社会システムデザイン』が、そして、ロボットや自動走行などによって人の可能性を広げるためには『コンピュータ工学』が必要となります。

ソフトウェア情報学

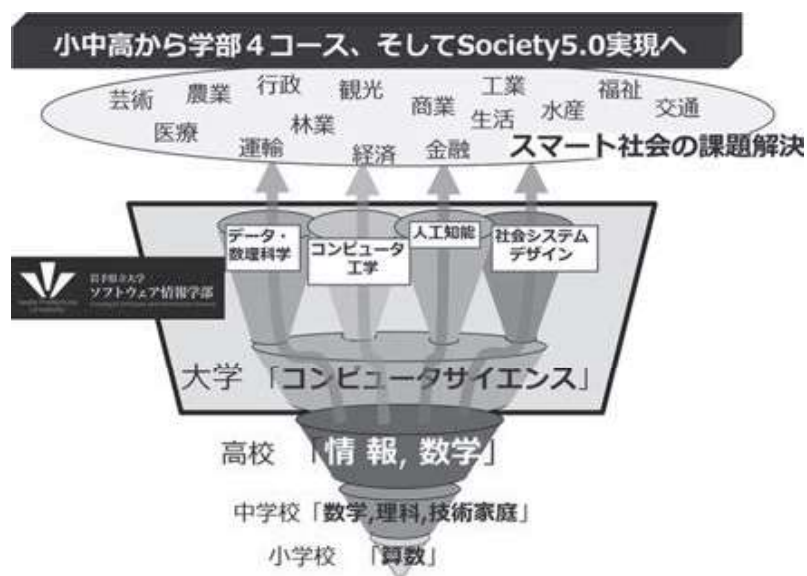


図1 平成 31 年度からの学部の 4 コース

部では、図1に示すように4つのコースを平成31年度からスタートします[2]。これら4つのコースでの学修では、「小学校『算数』」、「中学校『数学、理科、技術家庭』」、「高校『情報、数学』」の各教科を通じて育まれる基礎学力が重要です。つまり、大学でコンピュータサイエンスを学ぶための基礎が、小中高での学びです。学部入学後は、コンピュータサイエンスの基礎を学んだあと、4つのコースのいずれかを専攻し、卒業後に「スマート社会」実現のための課題解決に取り組みます。

「スマート社会」、この言葉の「スマート」とは何のことでしょうか。どんな意味で使われているのでしょうか。スマートがついた言葉はたくさんあります。例えば、スマートフォン、スマートカード、スマート家電などが挙げられます。なぜ、これらには「スマート」という言葉がついているのでしょうか。これらに共通なこと、どれにも含まれていることは何でしょうか？

「スマート」を辞書¹で調べると「コンピュータ制御の」が載っています。「スマート」とは、「コンピュータに制御されている」あるいは「コンピュータが内蔵されている」モノにつけられる言葉です。つまり、スマート社会は、あらゆるものにコンピュータが内蔵され、コンピュータによってコントロールされる社会です。では、そのコンピュータが処理しているモノは何でしょうか。それは「情報」です。スマート社会のもとでは、「新たな価値」や「必要な情報」がコンピュータによって生み出されます。では、この「情報」とは何でしょうか。以下では、「情報」についてお話しします。

2 「情報」を科学する

次のクイズを考えてみましょう。「ふ」から始まる言葉を完成させる問

¹ 『ジーニアス英和辞典 第5版』大修館書店（2014）より

題です。

ふ○○○○○○○○

各○にはひらがなが一文字ずつあてはまります。このままだと、何があてはまるか、検討もつきませんね。例えば、ヒントとして、

短歌、 石川啄木、 一握の砂

が与えられたとすれば、最初の5文字は「ふるさとの」であると予想できます。さらに、「盛岡駅、上野駅」がヒントとして与えられれば、候補は次の2つに絞られます。

ふるさとのなまり²

ふるさとのやまに³

このように、ヒントによって「情報」がもたらされ、分からなかったことが分かるように、曖昧だったことが明確になります。

この「情報」を科学的に考察してみましよう。

クイズの例では、情報（ヒント）によって、「情報が伝えられる前と後」では、「分かった、気がついた」ことが変化しました。このときの変化の程度を図2のように計る方法について考えましよう。

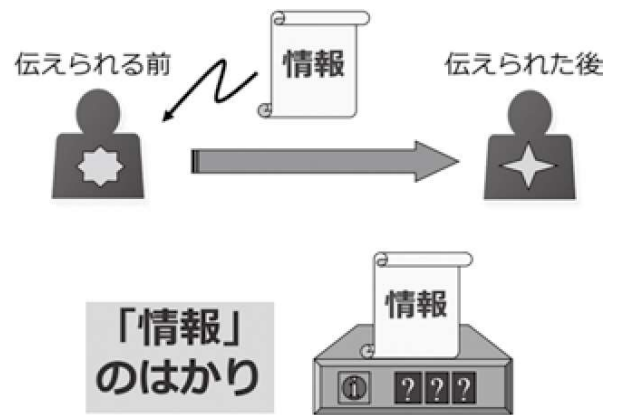


図2 情報のはかり

例えば、「役立たない情報、知っている情報」を「情報のはかり」で計ったら、「役立つ情報、知らない情報」の方が情報が大きい関係になるでしょう。

「役立つ情報」>「役立たない情報」

「知らない情報」>「知っている情報」

² 「ふるさとの訛（なまり） なつかし停車場の人ごみの中にそを聴きにゆく」（上野駅）

³ 「ふるさとの山に向ひて言ふことなしふるさとの山はありがたきかな」（盛岡駅）

このことを次の例題を用いながら説明します。図3に示すホテルでパーティーが開かれます。このホテルは、本館と新館に分かれており、パーティーが行える部屋はそれぞれの建物の1Fと2Fにあり、各階は北側と南側に分かれています。つまり、全部で候補が8部屋あります。このとき、会場を特定するために必要な「情報」は何でしょうか。

例：パーティー会場はどこ？

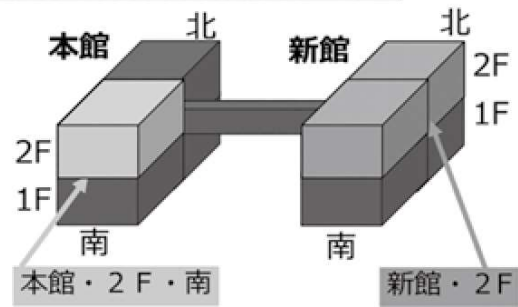


図3 例題：パーティー会場探し

例として次の3つの情報を考えます。

情報A：本館です

情報B：本館の2Fです

情報C：本館の2Fの南です

それぞれを「情報のはかり」で量ると、次の大小関係になるでしょう。

「本館の2Fの南です」 > 「本館の2Fです」 > 「本館です」

各情報が伝えられる「前」と「後」で変化したことは、候補となる部屋の数です。例えば、情報Aでは8から4、情報Bでは8から2、情報Cでは8から1です。このように、いずれも候補数が減少

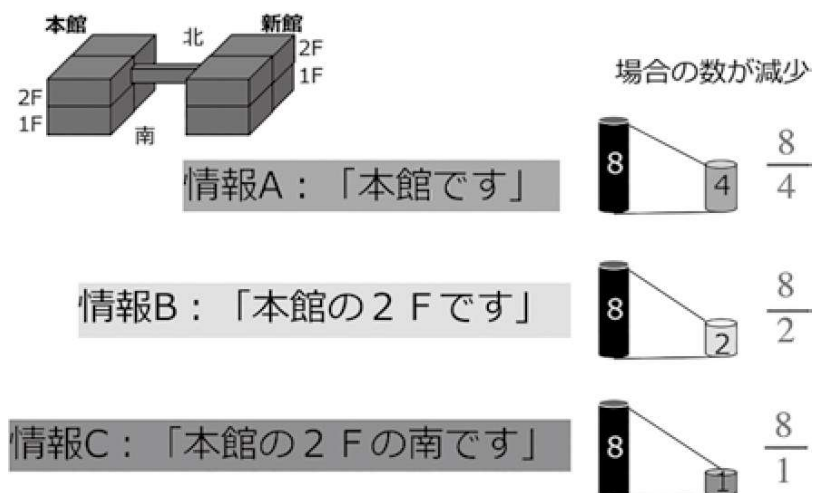


図4 情報による候補数の減少

しており、その減少の程度には情報による差がみられます。それを、図4のように「前」と「後」の比（商）で表してみましょう。これが「情報量」のもとになります。

情報による候補数の減少の様子をまとめたのが図5です。「本館です」によって8が4へ、さらに、「2Fです」によって、4から2へ、そして、「南です」によって2から1へ、候補がそれぞれ絞り込まれ、最終的に一つの部屋が決まります。

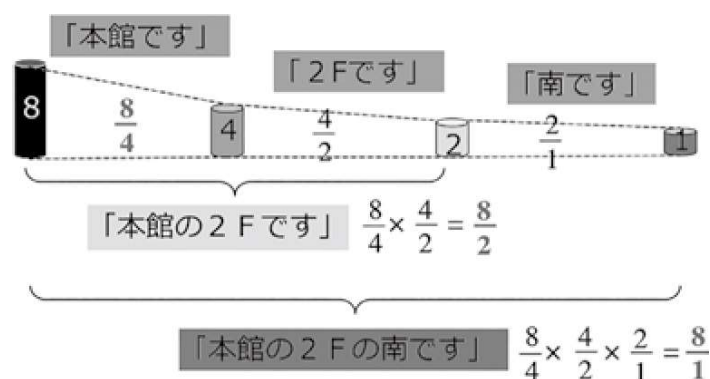


図5 情報による減少の割合

これら2つの情報を合わせた「本館2Fです」は、「本館です」と「2Fです」での減少の程度を表す、 $8/4$ と $4/2$ の積 $8/2$ と等しくなります。さらに、「本館の2Fの南です」は、 $8/4$ 、 $4/2$ 、 $2/1$ の積です。つまり、情報は加えることができ、そのことを数式で表せるのです。このときの数式では乗算なので、「情報の加法性」の言葉どおりに加算で表してみましょう。そのために、乗算を加算に変換する機能をもつ対数関数を利用します。対数関数のもとでは、例えば、積 $M \times N$ の対数 $\log_a M \times N$ は、 $\log_a M$ と $\log_a N$ の和と等しくなります（ a は底）。すなわち、

$$\log_a(2 \times 2 \times 2) = (\log_a 2) + (\log_a 2) + (\log_a 2)$$

が成り立ちます。以上の考察から、情報量を定めてみましょう。

伝えられた情報によって、候補の部屋が減少する程度が大きいほど、情報量が大きくなるよ

うに、図6のように、「伝えられる前の場合の数 A 」と「伝えられた後の場合の数 B 」の商 P を求めます。そして、 $P = 2^n$ を満たす n を「情報量

情 報	「前」の 場合の数 A	「後」の 場合の数 B	商 $A/B=P$	2^n	情報量 n
本館です	8	4	$8/4=2$	2^1	1
2Fです	8	4	$8/4=2$	2^1	1
本館の2Fです	8	2	$8/2=4$	2^2	2
本館の南です	8	2	$8/2=4$	2^2	2
本館の2Fの南です	8	1	$8/1=8$	2^3	3
本館または新館です	8	8	$8/8=1$	2^0	0

図6 情報の定量化

「 I 」と定めます。この情報量 I は P より次式で求められます。

$$I = \log_2 P \quad \text{[ビット]}$$

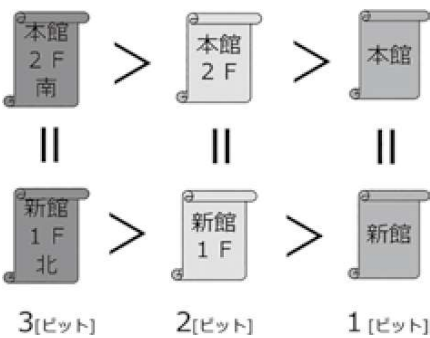


図 7 情報量の統計的性質

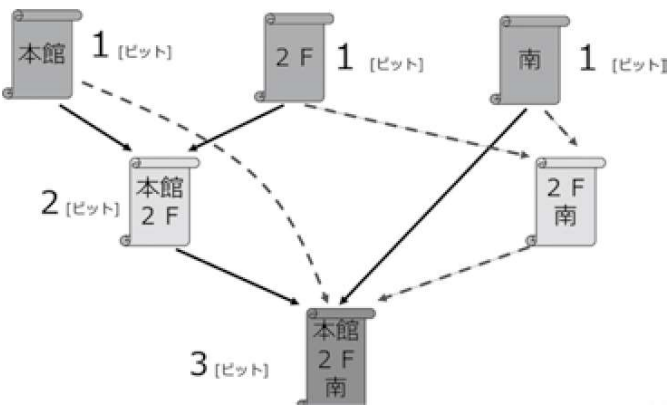


図 8 情報量の加法性

いま、定めた情報量の特性の 1 つ目は、図 7 のような統計的な性質です。すなわち、「本館です」と「新館です」は同じ情報量です。また、「本館の 2F です」と「新館の 1F です」も同じ情報量です。いずれも、場合の数が減少する程度が同じだからです。

情報量の特性の 2 つ目は、加法性です。図 8 のように、2 つの情報が加わったときの情報量は、それぞれがもつ情報量の和と同じになります。なお、2 つの情報を加える順番による情報量の変化はありません。つまり、「本館」であることが先に分かった場合と、「南側」であることが先に分かった場合による違いはありません。

3 情報量からエントロピーへ

「情報」は、質問をすることでも得られます。ここでは、質問の答えは、常に「はい」、「いいえ」とし、先のホテルの会場探しで、想定される質問の例として図 9 を考えます。各質問に

質 問	「はい」の可能性	「いいえ」の可能性
本館ですか？	4/8	4/8
2 F ですか？	4/8	4/8
本館の南ですか？	2/8	6/8
本館の 2 F ですか？	2/8	6/8
本館の 2 F の南ですか？	1/8	7/8

図 9 質問例

対して「はい、いいえ」、それぞれが回答される可能性、すなわち、確率はこの図のとおりです。回答「はい、いいえ」の可能性も考慮したときの情報量 H として、図10の式が提案されています。この式において、 P_Y は「はい」の回答、 P_N は「いいえ」の回答における値です。この式は、回答「はい」と「いいえ」によって伝えられる情報の平均値を求めるものであり、確率の分野における「期待値」に対応しており、「情報エントロピー」とも呼ばれています。

質問「本館ですか」の場合、

「はい」の可能性と、「いいえ」の可能性は、それぞれ $1/2$ なので、エントロピーは1ビットになります。さらに、図11のように「本館の2Fですか」に対して、「はい」の可能性は $2/8$ 、「いいえ」が「 $6/8$ 」より、エントロピーは0.81です。「本館の2Fの南ですか」の場合、「はい」の可能性は $1/8$ 、「いいえ」が「 $7/8$ 」より、エントロピーは0.54です。

情報エントロピー（以下、単にエントロピー）

$$H = \frac{1}{P_Y} \log_2 P_Y + \frac{1}{P_N} \log_2 P_N \quad [\text{ビット}]$$

質問	「はい」の可能性	商 $A/B=P$	「いいえ」の可能性	商 $A/B=P$
A 本館ですか？	4/8	8/4	4/8	8/4

$$H_A = \frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2 = 1 \quad [\text{ビット}]$$

図10 情報のエントロピー

質問	「はい」の可能性	商 $A/B=P$	「いいえ」の可能性	商 $A/B=P$	エントロピー [ビット]
本館ですか？	4/8	8/4	4/8	8/4	1.00
1Fですか？	4/8	8/4	4/8	8/4	1.00
新館の南ですか？	2/8	8/2	6/8	8/6	0.81
新館の2Fですか？	2/8	8/2	6/8	8/6	0.81
新館の2Fの南ですか？	1/8	8/1	7/8	8/7	0.54

図11 質問毎の情報のエントロピー

これらの例から分かるように、もっとも多くの情報量を獲得できる質問は、「本館ですか?」、「1Fですか?」に代表される候補を半分にする質問です。図12に示すように、1回の質問で、場合の数を半分にできれば、毎回1ビットの情報が得られ、それを3回繰り返すことで、会場を特

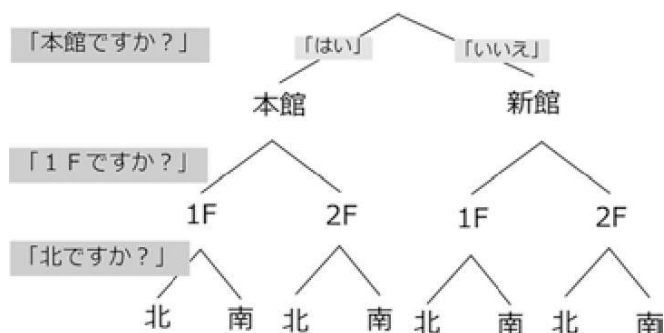


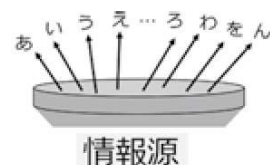
図12 質問による候補の絞り込み

定できます。8つの中から1つを選ぶには、3ビットの情報を得なければならず、「1ビットを得る質問を3回」行う必要があります。

4 生まれ出る情報

情報を発することができる資源、すなわち、「情報源」を考えます。この情報源からは図13のように「ひらがな」が発生されるとすれば、穴埋めクイズ「ふ○○○○○○○」

を解くことができます。情報源から各文字が出現する頻度（確率）が「等しいとき」と、「等しくないとき」とで分けてみましょう。濁音、半濁音なども含めたひらがな、合計



各文字の出現頻度
均一の場合
不均一の場合

情報源のエントロピー

$$H = \frac{1}{P_{あ}} \log_2 P_{あ} + \frac{1}{P_{い}} \log_2 P_{い} + \frac{1}{P_{う}} \log_2 P_{う} + \dots$$

図13 ひらがなを発生する情報源

81文字が、均一に発生する場合のエントロピー H は、図13の式より「約6.3ビット」となります。これに対して、日本語の活字資料をもとに、各文字の出現頻度を統計的に集計して得られた次表をもとにして算出したエントロピーは「約5ビット」です[3]。

文字	の	に	い	は	と	た	を	…	め	よ	…
頻度 (%)	9.3	9.3	9.3	3.3	3.0	2.6	2.0		1.2	0.9	

エントロピーが大きいほど、情報源から発生される情報は曖昧です。つまり、何が出てくるのか分からない、カオスの状態に近くなります。ところで、日本語に代表される自然言語では、次の例のように、隣接する文字によって、あてはまる文字が限定されます。

ふ○○との○○

ふ○さ○の○ま

ふるさ○のや○

このことから、直前の1文字を考慮してエントロピーを計算すると「約4.3ビット」、さらに、直前の2文字を考慮した場合のエントロピーはさらに減って「約2.9ビット」となります[3]。このように、実際に使われている文書中での文字の出現頻度を統計処理することで、情報源のエントロピーを減少させる、言い換えると、情報源の曖昧さを減少させることが可能です。

次に、情報源の考え方をを使って、穴埋めクイズを解いてみましょう。出現頻度の高い文字を使って、穴埋めしてみると、例えば、図14のような字句ができあがります。これは一例で、他にも考えられます。

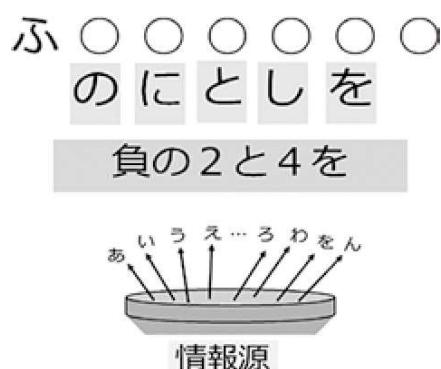


図14 情報源による解法



図15 情報源による作曲

さらに、図15のように、情報源から音が発生するとみなすと、作曲ができます。大量の楽譜から各音の出現頻度を求め、次々と音を並べることで曲ができあがります。この他にも、情報源を他のものにあてはめれば、いろいろな作品を作ることができるでしょう。

5 まとめ

これまでお話したことは、大学で学ぶ「コンピュータサイエンス」の基礎の一部にあたる内容です。学部の新コースでは、より高度かつ先端的な技術を学びます。

最後に今日のまとめを述べ、終えることにします。

- ・「情報」は統計的な性質をもとに計量可能

- ・情報の単位は「ビット」
- ・1 ビットは「2 つの中から 1 つを選ぶことができる知らせ」
- ・「多くの中から候補を絞り込める知らせ」の情報量は大
- ・複数個の知らせからなる「情報源」の平均情報量が「情報エントロピー」
- ・曖昧さが大（何が起こるか分からない）である情報源ほど、エントロピーは大きい

参考文献

- [1] 「Society 5.0」内閣府公式ページ
http://www8.cao.go.jp/cstp/society5_0/index.html
- [2] 「ソフトウェア情報学部のこれから」Web ページ
<http://www.soft.iwate-pu.ac.jp/outline/gakubufuture/>
- [3] 青柳 忠克：「情報とは一中・高校生からわかる情報の話」，産業図書（1992）
ISBN-13: 978-4782851333
- [4] 高岡 詠子：「シャノンの情報理論入門（ブルーバックス）」，講談社（2012）
ISBN-13: 978-4062577953
- [5] 川合 慧，萩谷 昌己：「基礎情報科学」，放送大学教育振興会（2009） ISBN-13: 978-4595139093