

AI/LLM(大規模言語モデル)の時代における機械・人間の 「理解モデル」の提案

水野 義之

関西外国語大学

ymizuno@kansai.ac.jp

Modeling of “Understanding” of Machines and Human in the AI/LLM Era

Yoshiyuki Mizuno

Kansai Gaidai Univ.

概要

この論考では、著者の提案する情報・知能モデルを基礎としてLLM（大規模言語モデル）を含む広義のAI（人工知能）の近年の発展が、情報教育に何をたらしかを考察する。特にその上で必要な認識、すなわち「理解する」とはどのようなことかに関して、「理解のモデル」を提案したい。ここで著者が提案する従前の情報・知能モデルには2つあり、1つは「情報のモデル」、もう一つは「知能のモデル」である。これらは有用かつ実用的で汎用性が高く、人間がAIを理解する上で必要と考えられる。しかし十分ではない。なぜなら情報と知能のモデルだけでは、人間が「考える」ことについて言及できないからである。そこでこの論考では「考える」こと自体の解明の第一歩として、新たに「理解する」とはどのようなことか考察する。「理解」とは「抽象化」とであるという認識に至ることを示す。

1 はじめに

この論考では、著者が提案する「情報の発展モデル」と「知能の発展モデル」を基礎として、LLM（大規模言語モデル）あるいは広義のAI（人工知能）の近年の発展が、情報教育に何をたらしかについて考える。

この問題を考察する上で必要な基礎的認識は、そもそも何かを「理解する」とはどのようなことか、という問題意識である。このため、この論考では、人間がこの問題を考えるために必要な「理解のモデル」を提案する。

なぜ、我々人間は「情報教育」の上で、「理解のモデル」を必要とするのか。それは、そもそも「教育」の目的は、人間（教育対象）自身が何かを「理解」したと思えるようにすること、にあると思われるからである。それにも関わらず、近年のLLMあるいは広義のAIの技術的進展は急激か

つ激越であり、人間教育に対する社会的影響と混乱は無視できないほど大きい。特に人間から見ると、機械が「何かを理解」しているかに見える点が、人間を困惑させる。LLMは著者自身にもそのような印象を与えることは事実である。そしてこのような事実こそが、社会における教育の意味理解に混乱をもたらしている。情報教育の関係者こそ、情報技術と人間の関係を最もよく理解しているはずである。従ってこの論考のような問題意識と課題の設定（「理解するとはどのようなことか」の理解）は、本来的に適切なはずである。

ここで以下の考察で参照するところの、著者の提案になる「情報の発展モデル」とは、「データ・情報・知識・知恵・統合」という「情報の5段階モデル」である[1]。また著者が提案する「知能の発展モデル」とは、「知性・理性・感性・悟性」という「知能の4成分モデル」のことを指す[4-6]。ちなみに、これらのモデルは著者独自のもの

のであるが、徐々に理解者は増えているようである。

前者「情報の発展モデル」は、情報の発展過程に対する評価に有用である。また後者「知能の発展モデル」は、知能とは何かを考える上で汎用性があることが特徴である。このためこれらの「情報と知能のモデル化」は、人間がAIを理解する上で、必須だと考えられるところの「人間知能」の理解を、人間自身が理解する上で必要な、モデルの一つであると考えられる。

しかしこれらは実は十分であるとは言えない。なぜなら、情報とこれを扱う知能の背後にあるところの、人間あるいは機械が「考えるとはどういうことか」について、これらのモデルだけでは、言及できていながらである。

そこで以下の論考では、「考えるとはどういうことか」という謎に迫る第一歩として、まず「理解する」とはどういうことかについて、解明することを目指す。

2 情報の発展モデル

著者が提唱しているところの「情報の発展モデル」[1]は、次のようなモデルを指す。通常の情報学の教科書[2,3]の冒頭に出てくる「情報モデル」は、「データ・情報・知識」を区別している。その先に「知恵」を措定する場合もある。

しかし著者はその先に「統合」を措定している。なぜなら、ちょうどバラバラでも情報は情報であるが、バラバラなものを知識とは呼ばないのと同じ理由で、バラバラでも知恵は知恵であるが、それを人間は何か一つのものとして理解したいと考えるためである。それを著者は「統合」と呼んでいる。

「統合」の例として芸術、社会、人間、自然全体、などを挙げる事ができる。

例えば物理学は、自然の統一的理解を目指す。それがこの統合の良い例である。ところで物理学の統一理論の理解が難易度が高くなるように、統合的なるものは一般に抽象度が高くなり、理解に困難を伴う。そこで統合の先には、分析的な理解増進（理解回復）のプロセスがあるはずで

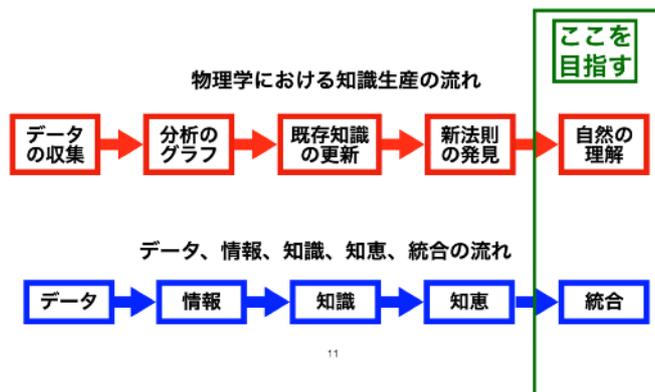


図1：物理学に見られる「統合」（統一的な自然理解）の流れ（上側）と、データ・情報・知識・知恵・統合における「統合」の流れ（下側）の類比あるいは対比と類似性。

「情報の発展モデル」（水野、2009）

出典：水野義之, 2009, 「情報社会における「情報」の発展モデル」
『日本社会情報学会第24回全国大会研究発表論文集』:184-187.

データ、情報、知識、知恵、芸術、科学

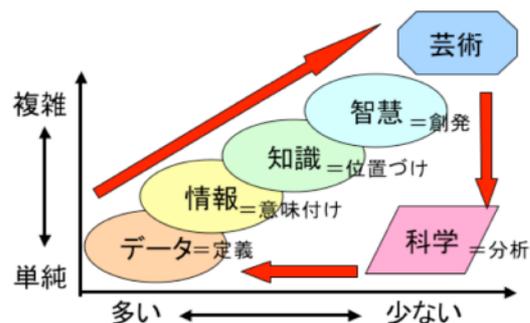


図2：情報の発展モデル[1]

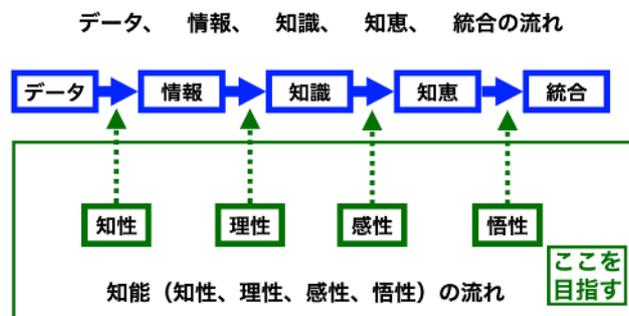


図3：「情報の発展モデル」（図5）における情報の5段階に対応して、人間の知能は、この情報発展を行うための4段階に分解できる。

ある。これを著者は「科学」と呼んでいる。

科学は新たなレベルで再度、データを生み出す。かくして、元に戻るがレベルは上がっているはずである。従ってこのような「情報の発展過程」は、螺旋的に永遠に発展を続けるはず、というモデルである（図2）。

3 知能の発展モデル

著者が提唱する「知能の発展モデル」は、次のようなものだ[4-6]。すなわち、2016年の「AI時代」の幕開け（囲碁AIのAlphaGoがトッププロ棋士に快勝した事象）と共に、気づいた疑問、AI（人工知能）はいいが「人間知能」について我々はわかっていないのではないか、という疑問に自問自答する中で、著者が発見した「知能のモデル」である。これは次のようなモデルだ。

まず情報の発展過程には5段階あった（データ・情報・知識・知恵・統合）。この各段階は、それぞれ異なるものである。従って、その形態に「変化」を起こす原因となるものがあるはずである。これを「人間知能」と考える。

このため、情報に5段階あることに対応して、「人間知能」には4成分あることが分かる。この4成分は「知性・理性・感性・悟性」という4成分に綺麗に対応する（図3）。

ここまで分かると、「データ・情報・知識・知恵・統合」の流れは、哲学分野で知られるところのイギリス流の経験論でいう、帰納法の推論の流れに対応することが分かる。これは知識の「内容としての科学」と呼ぶことも出来る（図4上の流れ）。

加えて、統合を理解するための分析としての科学は、図4下（左向きの流れ）に対応する。これを「方法としての科学」と呼ぶことが出来る。これは哲学分野で言われる、大陸の合理論つまり演繹的な分析の流れに対応することも分かる。

この2つのモデル化、情報モデルと知能モデルの両者、を合体させることで（図5）、情報と知能の中に、AIを、適切に位置付けることが、初

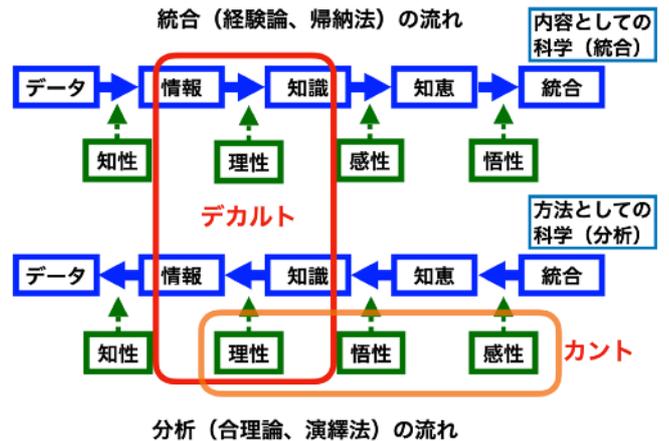


図4：論文[4-6]の「人間知能のモデル」と、歴史的に知られる知能モデル（デカルトの理性論とカントの認識論）の比較。

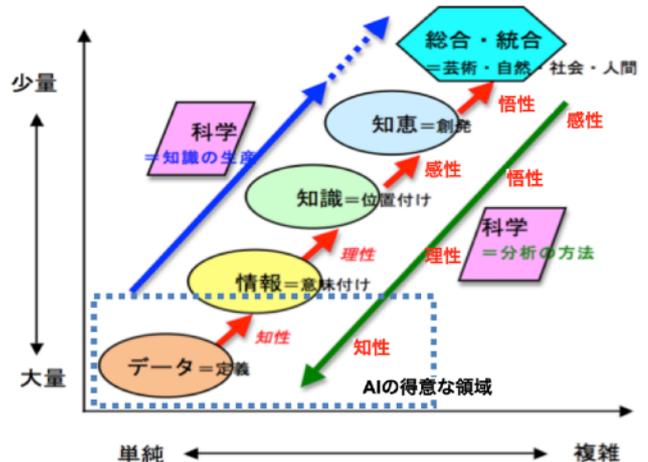


図5：画像認識など初期「AI（人工知能）の得意な領域」の位置付け。これは「情報の発展モデル」と「人間知能の発展モデル」のマップにおける左下の一角に対応する。

めて可能になったことが分かる。例えば、こういう具合である。

1) 画像認識などのAIは、データを情報に変換している。従ってこのAIは、人間の知性に対応することが分かる（つまり、人間の「知性」の一部を自動化して「置き換える」）。

2) また生成AIのように、質問文（プロンプト）に対応して文章を生成するAIは、情報から知識を構造化できる（図6）。従ってこのような生

成AIの知能は、人間の「理性」の一部に対応することが分かる（人間の「理性」と呼ばれる能力の一部を自動化して置き換える）。

4 新しい問題

ここまではわかったとしよう。しかし、問題はこれからである。すなわち不思議なのは、生成AIがあたかも「意識」を持っているかのように、我々は感じるのではないか、ということである。これはなぜか。当然ながら、AI（機械）が意識を持つとは誰も思わないだろう。しかし少なくともAIが「何かを理解」しているかのように感じる。これはなぜなのか。そういう疑問である。

ここでの考察の土俵を一旦、広げてみよう。例えば仏教の法相宗に、唯識あるいは唯識論・唯識思想と呼ばれる理論的理解がある。これは人間の無意識（唯識論では末那識[まなしき]あるいは阿頼耶識[あらいやしき]と呼ばれる）まで含んだ「意識の謎」を解明しようとする事が知られる。すなわち逆に意識の謎や心の謎まで、ここで相手にすると恐らく收拾がつかないと思われる。

そこで以下、本論考では、これらの「難問」を、次の5階層に分解して考える。つまりここでの考察の条件（対象）を狭める（限定する）ことで、却って考えやすくすることを選ぶ。

○これまでの考察

- 1) 情報の構造レイヤー（形態上の分類）：⇔「情報の発展モデル」
- 2) 知能の構造レイヤー（形態上の分類）：⇔「知能の発展モデル」

○本論文での新たな考察

- 3) 知能の機能レイヤー（機能上の分類）
- 4) 思考レイヤー（知能の機能を使う＝考える）
- 5) 意識（無意識）レイヤー

これを図にしてみよう。すると人間が「考える」ということの実体（内実）には、図7のような階層構造があるのではないか、と思われること

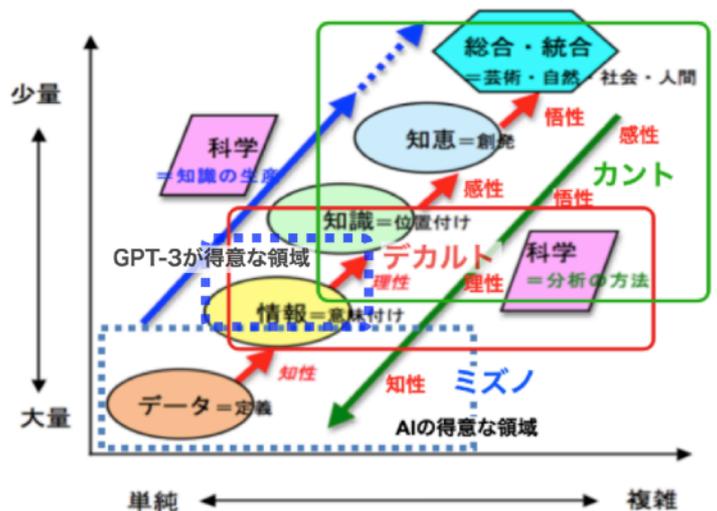


図6：図5に加えて、GPT-3が得意な領域（情報と知識の間。理性の一部）を、青の破線で追記した。

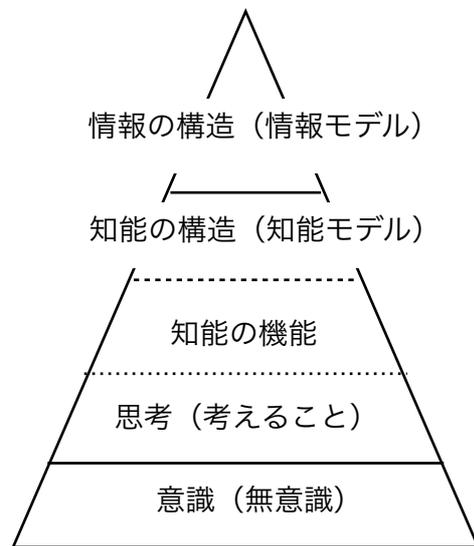


図7：人間が「考える」ということの実体（内実）の階層構造（作業仮説）。

がわかる（これは一種の「仮説」つまり「作業仮説」と解釈されたい）。

5 人間知能の機能レイヤーと抽象化

脳には構造と機能が、それぞれの場所場所に対応していることが知られている（ペンフィールドの脳地図）。これとのアナロジーで、我々はすでに「知能の」構造的あるいは形態上の分類を行ったのであるから（上記第3節）、これらにそ

れぞれ、何かの「機能が」対応すると考えることにする（仮定）。

さて、知性に対応する機能は何か。これは識別あるいは「分類」する作業に対応すると思われる。

理性に対応する機能は何であろうか。これは構造化つまり抽象化する作業に対応すると思われる。なぜなら構造に一つの名前をつけられるからである。

感性に対応する機能とは何か。これは創発と呼ばれる機能に対応するのではないか。

悟性に対応する機能は何か。これはより高度の抽象化に対応するのではないか。

悟性とは、ドイツ語でVerstand、英語ではunderstanding、つまり「理解する」ということである。しかし問題は、理解のプロセスはこの悟性だけではない、と思われることである。それでは、「理解する」とはどういうことなのか。

この「知能の4成分モデル」の「構造と機能の対応」において、最も重要なキーワードは「抽象化」である。

「抽象化」とは何か。抽象化とは、個別具体的に多種多様なものの中に、共通な何か（性質）を見出して、情報の量を減らすと同時に、普遍性・一般性を与えること。つまりそれに一つの「名前」を与える作業である。

抽象化の例を挙げる。例えば物理学の例で言うと、電磁気学のケースを考えると分かりやすい。電磁気学というのは、電気と磁気が織りなす多種多様で多彩な現象の総体を、総合的・統合的に研究する学問である。電磁気学は、パソコンやインターネットで誰でもいつもお世話になっている電気文明の基礎学問である。電磁気学は1つの体系化された学問である。この電磁気学の基礎方程式は、4本の個別の法則で出来ている。それぞれ名前がついている。曰く、電気現象（ガウスの法則）、磁気現象（磁場に関するガウスの法則）、電流の磁気作用（アンペールの法則）、磁場変化の電気作用（電磁誘導の法則）この4つである。しかしこれらは実は1つの連立方程式系として、1つの名前が付いている。これが電磁気学に関する「マクスウェル方程式」と呼ばれるも

のの実体である。それは1セットの連立偏微分方程式である。これが諸現象の抽象化の例である。

犬を考えると、この犬、あの犬、その犬種、それらの共通の1つの名前が犬である。人間でも同じである。人は全部違うのに、抽象的な1つの名前（人間）を与えている。このようにして抽象化で情報量を減らす。しかし抽象化によって、考える意味空間や思考空間を広げることができる。

理性という知能は、情報を知識にする段階で我々が行なっている構造化であると、私は言っている。しかし構造化して全体を1つの構造物にしているとも言える。すなわちバラバラの情報ではなく、1つのまとまった知識にしている。従ってここで我々が「理性」によって行なっている動作は、抽象化に他ならない。

すなわち理解すること（悟性）における作業の内容は、実は理性の働きにおいても、我々は行なっていることが判明するのである。この作業をここでは抽象化と名づける。

知恵を構造化して1つの統合にする例（悟性）の一つは、芸術である。一つの芸術作品は、さまざまな要素から成り立っている。それらの要素は、個別具体的な知恵の塊であるに違いない。たとえなこの文章表現、この光の描き方、このフレーズのサビの繰り返し、などである。しかし芸術は、それらが「1つのまとまった作品」になっている。従って統合の段階に見られるのも、理性と同じく抽象化であるが、より高度な抽象化であると考えられる。

以上の考察をまとめる。考えることとは、抽象化を行うことに他ならないことが分かった。ただし考えることは、抽象化に閉じるわけではない。例えば認知や認識あるいは識別（人間知能の中で知性だけを使用）も、我々は、考えていると感ずる。それは、この論考の言葉で言えば、知性も知能の一部であることで理解できる。

広義には知能（4成分ある）を使うことを、我々は広義に「考える」と呼ぶのであろう。また「理解する」とは、考えた結果、抽象化を行うことが出来た時。その時我々は理解したと感ずるのであろうと思われる。

何かを考えて理解する、と言う脳内現象を、このような「理解」するならば、理解という過程は、実際には、言い換え（自分の脳内における言語による抽象化）によって行うことができる。分かった（理解した）とは、そういうことである。

5 人工知能の機能レイヤーと抽象化

人間が何かを「理解すること」あるいは人間が「考えること」について、ここまでは分かっただけとしよう。

では、それは機械には出来ないであろうか。

もちろん、例えばChatGPTには、それが出来るように、人間には思われることこそが、人間を戸惑わせている。例えば、ChatGPTが得意な、長文の要約を考えてみよう。これは抽象化でなくてなんだろうか。例えば、何かの長文をChatGPTに提示し、それに適切な表題をつけよ、とプロンプトを与えれば、瞬時に、数秒で、適切な表題を出力するであろう。これは抽象化そのものだと思われる。これは人間が行うことと何が違うのか。あるいは同じなのか。ここには、2つの重要な違いがあると思われる。

違い1) 人間は無意識に、この理解作業と抽象化作業を行なっている。機械はプロンプトが無ければ何もしない。

違い2) 人間が行う理解過程と抽象化は（知能の4成分のどの機能でも同じである）、その人間の背景知識や価値観や体験や目的・目標等を基礎に行なわれているであろう。それがその人間らしさである。他方で機械は全世界のデジタルデータを「勉強」して行なっており、平均的である。もちろんこれに今後、仮にその人らしさを付け加えるができれば、それさえも機械化することは容易である。しかし現状では、出来ていない。しかもそれさえも「指示待ち」に他ならない。

人工知能にも、人間知能と類似の、知能の機能レイヤーがおそらく存在し、また抽象化の作業も可

能であると思われる。しかし上記のように、本質的に非常に大きな違いがあることも明らかになるのである。

6 まとめと結論・今後の課題

以上をまとめる。この論文では、まず次の4つのことを示した。第一に「情報の発展モデル」、すなわちデータ、情報、知識、知恵、統合、という一連の5つの（情報の）発展プロセスのモデルを改めて紹介した。第二に「科学」を、この「情報の発展モデル」の中で位置付けできることも改めて示した。第三に、人間の「知能」は4成分に分解できること（知性、理性、感性、悟性）を紹介した。第四に、「人間知能」を再理解する中で、AIの位置付けが明確になることを示した。例えば、画像認識AIは「知性」の一部を扱い、また生成AIは「理性」の一部を扱うことを示した。

次に本論文では、以上の議論をさらに抽象化することで、人間が「考える」ということ、「理解する」ということの実体（内実）には、階層構造があると思われる、という作業仮説を提示した。

このような一連の議論の一部はすでに、著者がこれまで発表した文献[1], [4-9]に考察の一部を記してきた。これに加えて本論文を含めることで初めて、「理解すること」あるいは「考えること」に対する理解が得られることが分かったと思われる。従って、AIあるいはLLMの最近の発展をも視野に含めつつ、このような「理解」を基礎として、今後の情報教育やデータ科学の一般教育に留まらず、おそらくあらゆる分野の教育をも対象として、それらをより良いものに改善することが可能であることが判明したと考えることができる。

今後の課題としては、より広範な哲学分野あるいは倫理学・宗教学分野の知見と、ここに記した論考との関係の解明、あるいは最近の発展が著しい認知の数学モデル（自由エネルギー原理）との関係の解明など、関連研究あるいは歴史的考察を参照するという課題の有用性と必要性を挙げたい。

このような一見迂遠な方法を通して初めて、今回提案した情報モデル・知能モデルと知能の機能のモデル（理解モデル、思考モデル）、をさらに拡張しつつ改善し、発展させることは可能であると思われる。なぜなら本論考で示した「理解」も「思考（考えること）」もその基礎には、それらを駆動するであろう意識（あるいは無意識）の問題が潜んでいると考えられるからである。ただしそのためには今後の研究が必要である。

9. 水野義之「AIっていったい誰なのよ：いまのAIは「アホ」なのか？—たかがAI, されどAI」, RAD-IT21 WEBマガジン, <https://rad-it21.com/ai/mizuno20180611/>, 2018年.

参考文献

1. 水野義之、「情報社会における「情報」の発展モデル」『日本社会情報学会第24回全国大会研究発表論文集』pp.184-187, 2009年.
2. 岩波講座「マルチメディア情報学」『〈1〉マルチメディア情報学の基礎』1999年10月.
3. 情報処理学会一般情報教育委員会(編集)『一般情報教育 (IT Text—一般教育シリーズ)』2020年9月.
4. Y.Mizuno, “Modeling of human intelligence applied to general education of informatics in AI era”, AXIES 2020.
5. 水野義之、「AI（人工知能）の理解を目的とする「人間知能」のモデル化提案と情報教育の改善」、現代社会研究科論集：京都女子大学大学院現代社会研究科紀要、第15号、2021年3月、pp.77-87.
6. 水野義之、「「人間知能」のモデル化におけるAI(人工知能)特に言語モデルGPT-3の位置付け」、AXIES 2021.
7. 水野義之「AI人材の情報倫理教育におけるインフォームド・コンセントを基盤とした能動的学修」『私立大学情報教育協会 教育イノベーション大会』, 2019年.
8. 水野義之「文科系大学におけるICT教育を再興する～アクティブ・ラーニング(AL)から人工知能(AI)の時代へ」、阿部勘一編著『ICT教育再考～文科系大学におけるICT教育の現状と課題』、noa出版, 2020年.