

学位論文題名

# 画像に対する発話を対象とした 概念獲得システムに関する研究

## 学位論文内容の要旨

近年、ペットロボットやコミュニケーションロボット、掃除ロボットなど、さまざまな家庭用ロボットが開発されている。しかし、現状ではそれらが一般家庭に広く普及しているとは言い難い。ホームロボットの普及を妨げている原因の1つは、ユーザのロボットに対する過剰な期待だと考えられる。特に日本においては「ロボットはなんでもできる」というイメージが先行しているため、現実との差が著しい。一方で、ユーザの期待に応えるためにロボットを高機能化していくと、ロボットを操作するユーザ側の負担が大きくなるというジレンマも抱えている。今後の家庭用ロボットは、人間と同等の対話処理能力を有することが必須となるものと考えられる。

人間の言語活動は、対話の中から言語を獲得する、記憶に基づいた対話を行う、質問に対して論理的な応答を行う、ユーモアや皮肉を理解する、環境や相手に適した言葉を選択する、感情を表現する、などの様々な能力から成り立っている。それらの能力のうち、著者は、自然言語による対話を通じて実世界の中で言語を獲得するシステムの実現を目指している。研究の第一歩として、ユーザの画像に対する発話から名詞概念(画像に対するラベル)を獲得していくシステムである SINCA(System for Noun Concepts Acquisition from utterances about Image) の構築・提案を行った。

これまでも様々な用途の対話システムが開発されてきたが、それらはすべて大人の言語処理能力をモデルとして構築されている。しかし、人間の言語能力は極めて複雑であるため、このようなアプローチでは汎用的な対話システムに到達することは非常に困難だと考えられる。著者は、人が言語を獲得する能力の実現が人と同等の対話能力を計算機上で実現するための手がかりになるのではないかと考えている。本論文では、認知科学における言語獲得に関する研究から得られた知見を参考にしながら、実際にコンピュータ上に言語獲得システムを構築して行った研究について述べる。

SINCA は、あらゆる事物はそれに対応するラベルを持っているという知識を備えているが、成人が持っているような語彙や文法の知識は与えられていない。そのような前提知識に基づいたアルゴリズムや、入力を限定せずユーザが撮影した画像とそれに対する自由な発話といった実データを対象とする点、音声や画像といったマルチメディアに対応している点に本研究の独創性がある。

SINCA の有効性については、様々な観点から評価し、実証を行った。

第一に、SINCA の重要なアルゴリズムの一つであるラベル獲得ルールを評価するために、三つの実験を行った。ラベル獲得ルールとは、名詞概念をより効率的に獲得するために、ユーザの入力から生成されるテンプレートである。一つ目の実験は、名詞概念を獲得するために必要な入力回数をラベ

ル獲得ルールを利用する場合と利用しない場合で比較し、ラベル獲得ルールが名詞概念獲得に与える効果の評価を行うものである。実験の結果、ラベル獲得ルールを利用することで一つのラベルを獲得するために必要な平均入力回数が 7.0 回から 5.6 回に減少し、より効率的な名詞概念獲得が可能になることが明らかになった。二つ目の実験は、Google による検索ヒット件数を用いて各ルールの一般性を測定するものである。実験の結果、Google の検索ヒット件数が 10,000 件以下のルールは、一般性が低く適用されにくいと判断でき、淘汰することが可能であることが明らかとなった。三つ目の実験は、World Wide Web 上に存在する日本語文書をコーパスとみなし、SINCA が生成するラベル獲得ルールの妥当性を調査する性能評価実験である。実験の結果、全体の約 26% のルールが 80% 以上の有効率 (Web 検索で得られたフレーズに、ラベル獲得ルールの変数部分に名詞を挿入した表現が含まれる割合) で名詞の抽出が可能であることが明らかになった。

第二に、ユーザが実際に SINCA を使用したときに受ける印象を評価する実験を行った。人間の心理や行動に関する調査分析技法として広く用いられている評定尺度法を用いた評価において、全ての項目で中間点である 4.0 ポイントを上回る結果が得られ、SINCA が幅広いユーザに好印象を与えることを確認した。また、SINCA と ELIZA (キーワードと簡単なテンプレートを用いて応答を生成し、対話を行うシステム) の比較を行った。その結果、SINCA の評点は ELIZA と比較して平均 1.0 ポイント以上上回っており、SINCA はキーワードやテンプレートを用いた簡単な対話システムよりも好印象を与えられることが明らかとなった。

第三に、SINCA にどのような入力を行うのが望ましいのかを検証するために、幼児の日常生活を収録したビデオデータを書き起こすことによって作成した独自の話し言葉データを入力文として用いて名詞概念を獲得する実験を行った。このビデオデータは、幼児のいる家庭にビデオカメラを設置し、大人と幼児の間で行われる日常的な会話を収録するという方法で収集されたものである。撮影対象となった家族は、2 歳 7 か月の男児、12 か月の女児 (ビデオ撮影開始時の年齢) とその両親の 4 人で構成されている。実験の結果、ビデオデータから収集した入力文から名詞概念を獲得することが可能であることが明らかになった。また、アンケートによって収集した書き言葉の入力文を用いた予備実験の結果と比較すると、ビデオデータから収集した入力文を用いた場合の方が効率的に名詞概念を獲得することが可能であることが明らかになった。これは、ビデオデータによる入力文には一語発話が多い (全体の 18.6%)、助詞の欠落が多い (全体の 42.6%)、表現に多様性がある (ラベルの直後の表現はアンケートの場合の 1.9 倍)、などの特徴があるためであると考えられる。これらのことから、幼児が日常生活の中で聞いている大人の発話は SINCA への入力として適切であることが示された。

SINCA で用いた名詞概念獲得手法は、文法知識などを用いていないため、他の品詞の獲得にも応用が可能であると考えられる。そこで、SINCA の名詞概念獲得アルゴリズムと同様に、語彙や文法の知識を持たない状態から色を表す形容詞を獲得する手法の提案を行った。SINCA にその手法を適用し、評価実験を行った結果、提案したアルゴリズムによって入力文から色を表す形容詞を抽出することが可能であることが明らかになった。現状では色を表す形容詞に獲得対象を限定しているが、画像処理部の強化などによって、幅広い形容詞を獲得することも可能であると考えられる。

# 学位論文審査の要旨

主 査 教 授 荒 木 健 治  
副 査 教 授 長 谷 山 美 紀  
副 査 教 授 山 本 強

学 位 論 文 題 名

## 画像に対する発話を対象とした 概念獲得システムに関する研究

近年、ペットロボットやコミュニケーションロボット、掃除ロボットなど、さまざまな家庭用ロボットが開発されている。しかし、現状ではそれらが一般家庭に広く普及しているとは言い難い。ホームロボットの普及を妨げている原因の1つは、ユーザのロボットに対する過剰な期待だと考えられる。特に日本においては「ロボットはなんでもできる」というイメージが先行しているため、現実との差が著しい。一方で、ユーザの期待に応えるためにロボットを高機能化していくと、ロボットを操作するユーザ側の負担が大きくなるというジレンマも抱えている。今後の家庭用ロボットは、人間と同等の対話処理能力を有することが必須となるものと考えられる。人間の言語活動は、対話の中から言語を獲得する、記憶に基づいた対話を行う、質問に対して論理的な応答を行う、ユーモアや皮肉を理解する、環境や相手に適した言葉を選択する、感情を表現する、などの様々な能力から成り立っている。それらの能力のうち、著者は、自然言語による対話を通じて実世界の中で言語を獲得するシステムの実現を目指している。研究の第一歩として、ユーザの画像に対する発話から名詞概念(画像に対するラベル)を獲得していくシステムである SINCA(System for Noun Concepts Acquisition from utterances about Image) の構築・提案を行った。

これまでも様々な用途の対話システムが開発されてきたが、それらはすべて大人の言語処理能力をモデルとして構築されている。しかし、人間の言語能力は極めて複雑であるため、このようなアプローチでは汎用的な対話システムに到達することは非常に困難だと考えられる。著者は、人が言語を獲得する能力の実現が人と同等の対話能力を計算機上で実現するための手がかりになるのではないかと考えている。本論文では、認知科学における言語獲得に関する研究から得られた知見を参考にしながら、実際にコンピュータ上に言語獲得システムを構築して行った研究について述べる。

SINCA は、あらゆる事物はそれに対応するラベルを持っているという知識を備えているが、成人が持っているような語彙や文法の知識は与えられていない。そのような前提知識に基づいたアルゴリズムや、入力を限定せずユーザが撮影した画像とそれに対する自由な発話といった実データを対象とする点、音声や画像といったマルチメディアに対応している点に本研究の獨創性がある。

SINCA の有効性については、様々な観点から評価し、実証を行った。

第一に、SINCA の重要なアルゴリズムの一つであるラベル獲得ルールを評価するために、三つの実験を行った。ラベル獲得ルールとは、名詞概念をより効率的に獲得するために、ユーザの入力から

生成されるテンプレートである。一つ目の実験は、名詞概念を獲得するために必要な入力回数を、ラベル獲得ルールを利用する場合と利用しない場合で比較し、ラベル獲得ルールが名詞概念獲得に与える効果の評価を行うものである。実験の結果、ラベル獲得ルールを利用することで一つのラベルを獲得するために必要な平均入力回数が 7.0 回から 5.6 回に減少し、より効率的な名詞概念獲得が可能になることが明らかになった。二つ目の実験は、Google による検索ヒット件数を用いて各ルールの一般性を測定するものである。実験の結果、Google の検索ヒット件数が 10,000 件以下のルールは、一般性が低く適用されにくいと判断でき、淘汰することが可能であることが明らかとなった。三つ目の実験は、World Wide Web 上に存在する日本語文書をコーパスとみなし、SINCA が生成するラベル獲得ルールの妥当性を調査する性能評価実験である。実験の結果、全体の約 26% のルールが 80% 以上の有効率 (Web 検索で得られたフレーズに、ラベル獲得ルールの変数部分に名詞を挿入した表現が含まれる割合) で名詞の抽出が可能であることが明らかになった。

第二に、ユーザが実際に SINCA を使用したときに受ける印象を評価する実験を行った。人間の心理や行動に関する調査分析技法として広く用いられている評定尺度法を用いた評価において、全ての項目で中間点である 4.0 ポイントを上回る結果が得られ、SINCA が幅広いユーザに好印象を与えることを確認した。また、SINCA と ELIZA (キーワードと簡単なテンプレートを用いて応答を生成し、対話を行うシステム) の比較を行った。その結果、SINCA の評点は ELIZA と比較して平均 1.0 ポイント以上上回っており、SINCA はキーワードやテンプレートを用いた簡単な対話システムよりも好印象を与えられることが明らかとなった。

第三に、SINCA にどのような入力を行うのが望ましいのかを検証するために、幼児の日常生活を収録したビデオデータを書き起こすことによって作成した独自の話し言葉データを入力文として用いて名詞概念を獲得する実験を行った。このビデオデータは、幼児のいる家庭にビデオカメラを設置し、大人と幼児の間で行われる日常的な会話を収録するという方法で収集されたものである。撮影対象となった家族は、2歳7か月の男児、12か月の女児 (ビデオ撮影開始時の年齢) とその両親の4人で構成されている。実験の結果、ビデオデータから収集した入力文から名詞概念を獲得することが可能であることが明らかになった。また、アンケートによって収集した書き言葉の入力文を用いた予備実験の結果と比較すると、ビデオデータから収集した入力文を用いた場合の方が効率的に名詞概念を獲得することが可能であることが明らかになった。これは、ビデオデータによる入力文には一語発話が多い (全体の 18.6%)、助詞の欠落が多い (全体の 42.6%)、表現に多様性がある (ラベルの直後の表現はアンケートの場合の 1.9 倍)、などの特徴があるためであると考えられる。これらのことから、幼児が日常生活の中で聞いている大人の発話は SINCA への入力として適切であることが示された。

SINCA で用いた名詞概念獲得手法は、文法知識などを用いていないため、他の品詞の獲得にも応用が可能であると考えられる。そこで、SINCA の名詞概念獲得アルゴリズムと同様に、語彙や文法の知識を持たない状態から色を表す形容詞を獲得する手法の提案を行った。SINCA にその手法を適用し、評価実験を行った結果、提案したアルゴリズムによって入力文から色を表す形容詞を抽出することが可能であることが明らかになった。現状では色を表す形容詞に獲得対象を限定しているが、画像処理部の強化などによって、幅広い形容詞を獲得することも可能であると考えられる。

これを要するに、著者は、実世界で動作可能な概念獲得手法を提案し、言語獲得について有効な計算処理機構に関する新知見を得たものであり、自然言語処理工学及び人工知能工学の発展に貢献するところ大なるものがある。よって著者は、北海道大学博士 (情報科学) の学位を授与される資格あるものと認める。