

# 反転学習とシステム開発演習を活用する テキストアノテーション

松吉 俊<sup>1,a)</sup>

**概要:** 本研究では、テキストアノテーションに関する新しいパラダイムを提案する。従来のテキストアノテーション作業では、一般に、アノテーション指針の作成者自身がアノテーション完了まで多くの作業者を管理する。本論文では、反転授業(反転学習)を取り入れ、かつ、自然言語処理システムの開発実習にも取り組ませることにより、指針作成者が不在の状況であっても質の良いアノテーションを実現する枠組みを提案する。具体的な実践例として、「発話内の語句が曖昧であった時に人間らしく応答する対話システム」に関するテキストアノテーション演習授業について報告する。このアノテーション作業では、従来の「読み手によるアノテーション」ではなく、「書き手によるアノテーション」を実施した。書き手アノテーションのラベルは、対話システム等の人間らしさを自動評価するための1つの手段として利用できる可能性があることを述べる。

## A Novel Scheme for Text Annotation based on Flipped Learning and Project-Based Learning

SUGURU MATSUYOSHI<sup>1,a)</sup>

### 1. はじめに

現在、自然言語処理の研究では、テキストデータを利用することが主流となっている。機械学習手法に基づいて自然言語処理ツールを開発する場合、正解ラベルが付与された大規模テキストデータを利用することが多い。教師なし学習を採用する場合であっても、研究のすべての段階において、正解ラベルが付与されたデータを全く利用しないわけではなく、システムの性能評価の段階においてはそのようなデータを必要とする。

対象となるデータに対して、正解となるラベル、もしくは、それに類する情報を付けることを、**アノテーション**と呼ぶ。画像に何が写っているかを見て、その情報をテキストで表現することは、画像アノテーションの1つの例である。楽曲にコード進行の情報を付与することは、音楽アノテーションの1つの例である。文章のジャンルや難易度を

推定し、その情報をテキストで表現することは、**テキストアノテーション**の1つの例である。本研究では、テキストアノテーションに関する枠組みを研究対象とする。

伝統的なテキストアノテーション作業では、少数の管理者のもとに少数もしくは多数の作業者が集まり、アノテーションの指針に従ってテキストにラベルを付与していく。近年、この枠組みが持つ欠点に悩まされない枠組みとして、クラウドソーシングに基づくアノテーションやゲーミフィケーションに基づくアノテーションが提案され、多くの研究で利用されている。

本論文では、これらと異なる特徴を持つ新しいアノテーションの枠組みを提案する。提案する枠組みは、唯一の正しい枠組みという性質のものではない。タスクや状況に応じて、既存の手法を採用するか、この提案手法を採用するか、選択することが推奨される類のものである。提案手法は、既存の手法を補う特徴も持っているため、既存の手法と提案手法を組み合わせることも可能である。

提案手法の実践において、従来の「読み手によるアノ

<sup>1</sup> 電気通信大学  
The University of Electro-Communications  
<sup>a)</sup> matuyosi@uec.ac.jp

表 1 各アノテーション枠組みの特徴

		伝統的枠組み	クラウドソーシング	ゲーミフィケーション	FL+PBL (提案)
目的	主	アノテーション	アノテーション	アノテーション	履修者の学習
	副	-	-	-	アノテーション
経費	作業者雇用	高い	安い	なし	なし
指針設計	専門性	言語学	言語学	言語学, ゲーム設計	言語学, 教育学
	言語現象	制限なし	制限される	制限される	制限なし
	難易度	タスク依存	慣れにより低減	高	高
	管理者	兼ねる	兼ねる	不要	必須ではない
管理	専門性	不問	不問	管理不要	プログラミング指導
	獲得スキル	なし	タスク分割	-	プログラミング指導
作業	専門性	言語学	不問	不問	不問
	動機	賃金	賃金	楽しみ	単位取得, 興味
	不正の恐れ	なし	あり	あり	あり
	獲得スキル	アノテーション	なし	ゲームのコツ	アノテーション NLP の初歩
	フィードバック	ほぼ即時	即時	不可	次年度
成果物	納期	定めない	即納	制御不可	厳守
	質	高い	不均一	初期は高い	比較的高い
	量	調整可能	多い	予測不可	少ない
後日増量	問題点	コスト	特になし	制御不可	特になし
長所		堅実, 汎用的	即時性	経費不要, 初期の質	汎用的, 教育的
短所		コスト	対象の制限, 質	対象の制限	設計

ーション」ではなく、「書き手によるアノテーション」を実施した。書き手アノテーションのラベルは、対話システム等の人間らしさを自動評価するための1つの手段として利用できる可能性があることもこの論文で述べる。

本論文は、以下のように構成される。まず、2章において、既存の手法と比較しながら、提案手法について説明する。次に、3章で、提案する枠組みの具体的な実践例を紹介する。ここでは、「発話内の語句が曖昧であった時に人間らしく応答する対話システム」に関するテキストアノテーション演習授業の設計について述べる。続く4章で、演習授業によって得られたアノテーション済みテキストデータについて分析し、その利用方法を考察する。5章はまとめである。

## 2. アノテーション枠組み

### 2.1 準備

アノテーションの新しい枠組みに関する研究は、当然ながら、テキストアノテーションに関するものに限られない。それゆえ、参考文献は多分野にわたり数多く存在する。

既存のアノテーションの枠組みを説明するにあたり、本論文では、文献 [4] を参照する。この文献は、機械学習のためのテキストアノテーションについて解説する本である。文献の大部分は、伝統的なテキストアノテーションについての説明であるが、最終章 (12 章) に新しいテキストアノテーション枠組みについて簡潔にまとめられている。

本論文では、テキストアノテーションに関して、以下の

ように用語を定義する。

**言語現象** テキストデータにアノテーションする目的となる言語学的事項。例えば、語分割、係り受け関係、発話意図、文章ジャンルなど

**アノテーション指針** テキストデータにどのようにアノテーションを施すべきかに関する指針。アノテーション作業マニュアルと同義

**ラベル** アノテーション作業において実際に付与する情報。選択式でなく筆記式で文を書くアノテーションに関しても、便宜上、本論文ではラベルと呼ぶ。例えば、提示された発話に対してその返答文を記述する場合であっても、便宜上、その返答文のこともラベルと呼ぶ

**正解ラベル** 人手で付与した情報。誤りを含む可能性がある自動付与ラベルと区別する場合に用いる

**ラベル体系** 選択可能なラベルの集合。一般に、これが複雑である場合、アノテーション作業の難易度が上がる

**指針設計者** ラベル体系を定め、アノテーション指針を設計する人物

**作業者** アノテーション指針に基づき、実際にアノテーションする人物

**管理者** アノテーション指針に基づき、作業者を管理する人物

**利用者** アノテーション済みのテキストデータを利用する人物、もしくは、計算機

**履修者** 授業を受ける人物

**TA** ティーチングアシスタント。授業の補佐をする人物。

授業の設計はしない

**学習** 人間が行う学習、計算機が行う学習については、本論文では「機械学習」と「機械」を付けて呼ぶ

**アノテーション結果の質** 指針設計者が想定したとおりにラベルが付与されている場合、質が高いと表現する。複数作業によるアノテーション結果の一致度が高いからといって、必ずしもその質が高いとは限らないことに注意する必要がある

## 2.2 既存の枠組み

既存の代表的なアノテーションの枠組みとして、次の3つについて説明する。

- 伝統的なテキストアノテーション
- クラウドソーシングに基づくアノテーション
- ゲーミフィケーションに基づくアノテーション

表1にこれらの特徴<sup>\*1</sup>についてまとめる。

### 2.2.1 伝統的なテキストアノテーション

20世紀半ばから、言語学の分野で用いられている枠組みであり、現在は、自然言語処理の分野においても代表的なアノテーション方法である。

伝統的なテキストアノテーション作業では、まず、対象の言語現象に詳しい人物が指針設計者となる。指針設計者は、ラベル体系を定め、アノテーション指針を設計する。アノテーション指針の草稿が完成すると、実際に作業員にアノテーション作業を依頼する。ラベル体系とアノテーション指針が固まっている場合、指針設計者と管理者が異なる場合がある。しかしながら、ラベル体系やアノテーション指針の不備が、アノテーション作業からのフィードバックによって指摘されることが多い。このフィードバックを、アノテーション指針に迅速に反映させるために、通常、指針設計者が管理者を兼ねる。

この枠組みの一番の長所は、どのような言語現象でも対象とすることができることである。他の長所としては、**アノテーション作業からのフィードバック**を受けることにより、動的にラベル体系を改善できることがある。例えば、指針設計者が見落としていたラベル間の関係を、作業員が指摘することにより、指針設計者がラベル体系をより良いものにすることができる。また、アノテーション結果の質が低い場合に、ラベル体系を見直すことにより、アノテーション結果の質を速やかに向上させることが可能である。

この枠組みの短所は、概して、アノテーションが完了するまでに、費用と時間がかかることである。多くの場合、指針設計者や作業員は、専門的な知識を持った人物であり、その雇用に相応のコストがかかる。ラベル体系とアノテーション指針を更新しながら、質の高いアノテーション済み

<sup>\*1</sup> 本節および表1で述べることは一般論である。それゆえ、対象とする言語現象やアノテーション環境によっては、当てはまらないことがある。

データの構築を目指すため、作業完了までに時間もかかる。

### 2.2.2 クラウドソーシングに基づくアノテーション

伝統的な枠組みの短所に悩まされない方法として、クラウドソーシングを利用する方法が提案された。

**クラウドソーシング**とは、インターネット上で不特定多数の人物に仕事を依頼するシステムのことであり、このシステムを簡単に利用できる環境を提供するサービスとして、Amazon Mechanical Turk<sup>\*2</sup>やランサーズ<sup>\*3</sup>が有名である。

クラウドソーシングに基づくアノテーションは、まずは、画像処理の分野 [3], [5] で採用された。その後、その長所が他分野でも認識され、自然言語処理の分野においても、テキストアノテーションにおける選択肢の1つとして、その立場が確立されるに至った。クラウドソーシングに基づくアノテーションでは、作業員は、インターネットの向こうに存在する不特定多数の人物となる。

この枠組みの長所は、アノテーション完了までに、それほど多くの費用と時間を要しないことである。不特定多数に投げたアノテーション作業は、即日のうちに結果が返ってくることが多い。仕事内容に応じて設定すべき相場はほぼ決まっており、妥当な金額で安く仕事を発注することができる。

この枠組みの一番の短所は、依頼できる仕事内容がとても単純なものに限られることが多いことである。複雑な仕事である場合、その仕事を引き受けてくれる作業員がいなかったり、引き受けてくれたとしてもアノテーション結果の質が低かったりする。それゆえ、管理者には、アノテーション指針をじっくりと読み込み、アノテーション作業を丁寧に細分化することが求められる。この細分化は、経験と勘に基づく作業であり、たいいていの場合、不特定多数との複数回のやりとりにより試行錯誤する必要がある。他の短所としては、アノテーション結果の質は均一でないことがあげられる。丁寧にラベル付けをしてくれる作業員がいる一方で、すべて同じラベルを選択し、短時間で作業を終える作業員もいる。それゆえ、毎回、アノテーション結果の妥当性を評価することが必須となる。

### 2.2.3 ゲーミフィケーションに基づくアノテーション

アノテーション作業において、作業員に労働感ではなく楽しさを感じてもらおう方法が提案された。

ゲーム形式を利用するアノテーションには、次の2種類がある。

**ゲーミフィケーション** アノテーション作業自体をゲーム形式で行う

**Games with a Purpose (GWAP)** 作業員は、アノテーションしていることに気づかずに、ゲームに興じる。その裏で管理者がアノテーション結果を得る。本来これらは異なるものであるが、これらを総称してゲー

<sup>\*2</sup> <https://www.mturk.com/mturk/welcome>

<sup>\*3</sup> <http://www.lancers.jp>

ミフィケーションと呼ぶことにする。

この枠組みに関して、画像処理の分野では ESP game [7] が有名である。音声言語処理や自然言語処理の分野では、Siri<sup>\*4</sup>や Cortana<sup>\*5</sup> などが、認識率向上のための音声データ獲得や検索文字列の収集にこの枠組みを利用している。GWAP では、作業者は、インターネットの向こうに存在する不特定多数の人物である。現在稼働している多くのシステムでは、技術的には可能であるが、プライバシーポリシーにおいて個人情報特定しないと明記されていることが多い。狭義のゲーミフィケーションに基づく枠組みに関しては、作業者は、特定少数の場合も不特定多数の場合もありうる。

GWAP は、クラウドソーシングの 1 種であるので、2.2.2 節で述べた長所と短所がそのまま当てはまる。

広義のゲーミフィケーションに基づく枠組みの長所は、作業者が積極的にアノテーションに貢献してくれることである。

この枠組みの一番の短所は、ゲーム形式として実現するための設計がとても難しいということである。適度な難易度を持っており、飽きずに無理なく続けることができる仕様を考えるのは容易ではない。他の短所としては、ゲーム内のスコアを上げることに専念し、不正を働く作業者が現れやすいことや、GWAP の場合、ゲームに興じてくれる作業者の数が事前に把握できないことが挙げられる。

## 2.3 提案する枠組み

本論文では、テキストアノテーションに関する新しいパラダイムを提案する。この枠組みでは、近年、大学教育などにおいて積極的に推進されている次の 2 つの授業形態を活用する。

- 反転学習 (= 反転授業)
- 問題解決型の学習

以下、これらの方法について紹介したのちに、提案する枠組みについて説明する。

### 2.3.1 反転学習 (FL)

反転学習 (Flipped Learning; FL) とは、次のような授業形態 [6] である。

- 授業資料や説明動画を授業前に学生に公開する
- 履修者は事前にしっかり予習する
- 正規の授業時間は、議論や質疑、演習に当てる

通常の授業形態では、正規の授業時間内で説明を行い、帰宅後に復習や宿題 (演習) をさせる。正規授業時間の使用方法が反転していることから、反転学習と呼ばれる。反転学習を教師の立場から見て、反転授業 (Flipped Classroom) と呼ぶことがあるが、同じものを指す。

履修者が持つ前提知識や理解度は個々で異なる。予習し

た内容に基づいて、正規の授業時間中に履修者間で議論や質問をさせることにより、各履修者が内容の理解を深めることを目指す。正規の授業時間中に、演習時間を十分に取ることができるため、知識の定着や応用力の育成に期待が持てる教育方法である。

### 2.3.2 問題解決型の学習 (PBL)

問題解決型の学習 (Project-Based Learning; PBL) とは、次のような授業形態 [1] である。

- 授業に先立ち、テーマが与えられる
- 履修者は、自律的・主体的にこのテーマに取り組むことを求められる
- 履修者は、自らテーマに関する課題を見つけ、その解決を目指す
- チームで取り組むことが多い

PBL は、以下のような特徴を持っており、問題解決のための基礎能力を伸ばすことに期待が持てる教育方法である。

- プロジェクト期間が限定されている
- 手近なマニュアルがない
- トラブルがつきものである
- 自分 (たち) でスケジュールを決める必要がある

PBL にはトラブルがつきものであるため、実際の授業においては、TA や担当教員が、適宜、必要最低限のアドバイスをを行う。

### 2.3.3 FL+PBL を活用する枠組み

本研究では、反転学習を取り入れ、かつ、自然言語処理システムの開発実習にも取りこませることにより、指針設計者が不在の状況であっても質の良いテキストアノテーションを実現する枠組みを提案する。本研究は、言語学、教育学、情報学の 3 分野にわたる学際的研究である。表 1 の「FL+PBL (提案)」に、提案するテキストアノテーションの枠組みについてまとめる。

提案する枠組みを、準備段階と演習段階に分けて説明する。次の 3 章では、この枠組みの実践例を具体的に紹介する。

#### 準備段階

- (1) 対象の言語現象を定める
- (2) これに関する自然言語処理システムの簡易版を用意する
- (3) このシステムの開発に必要なような、アノテーション付きテキストデータ (アノテーション例) を少量用意する
- (4) 指針設計者が不在であっても良い程度に、授業内容を説明する資料を作成する
- (5) 最終課題として、次の 2 つを設定する
  - (a) アノテーション付きテキストデータの増量
  - (b) 簡易版システムの改良
- (6) 反転学習の予習に利用できるよう、以下の 3 つを事前に履修者に公開する

\*4 <https://www.apple.com/jp/ios/siri/>

\*5 <https://www.microsoft.com/ja-jp/windows/cortana>

授業資料, 簡易版システム, アノテーション例

### 演習段階 (正規の授業時間中)

(1) 授業資料の指示に従い, 締め切りまでに最終課題を達成するよう伝える

(2) TA が主にプログラミングに関する質問に答える

本枠組みの主目的は, 他の枠組みと明確に異なり, 履修者が自然言語処理に関する基礎技術を身につけることである。あえて補足するならば, 履修者に自然言語処理という分野が存在することを伝え, 「自然言語処理は楽しいものである」と興味を持ってもらうことである。それゆえに, 本枠組みでは, 至る所に教育上の配慮を凝らす必要が生じる。例えば, 実践においては, 以下のことに配慮しながら, 課題を設定する必要がある。

- プログラミングが得意でない履修者でも無理なくできる
- 留学生が不利にならない
- 履修者自らが最終課題の目標点を設定できる

(履修者が自身の判断で 60 点を目指したり, 100 点を  
目指したりできる)

大学の授業として実施する場合には, 次のことも重要である。

- 管理者として働く TA も, 演習授業を通してスキルを獲得できるようにする
- レポートの不正コピーを確実に発見できるような課題にする

本枠組みの長所は, 授業終了時にアノテーション済みテキストデータが得られるのみでなく, 自然言語処理技術者の観点から作業者が成長できる点\*6にある。他の長所としては, 一度作成された授業資料は, 一般公開されていれば, 他の機関においても再利用可能であり, 同様のアノテーション済みテキストデータが得られることが期待できることである。これは, 本枠組みが, 反転学習を採用しており, 指針設計者が不在の状況であってもアノテーション作業が進むよう設計されているからである。

本枠組みの短所は, 上記準備段階の (2), (3), (4) の設計が難しいことである。その理由は, これらの設計に, 言語学や自然言語処理の知識だけでなく, 教育に関する経験・知識も必要とされるからである。このような実践に関する研究報告も少なく, ノウハウが全く溜まっていないことも問題の 1 つである。本枠組みを用いた授業設計の敷居を下げるためにノウハウを蓄積することは, 今後の課題である。

表 1 の下部に示されるように, どの枠組みも長所と短所を持っている。新しくアノテーションを始める際には, これらの長所と短所を理解し, 適切なテキストアノテーション枠組みを選択することが推奨される。

\*6 演習中やレポートの採点中に, 自然言語処理研究に関して優れた人材を発見できることもある。

表 2 対話システムの演習に関して決定すべき項目

項目	選択肢
参加人数	2人, 3人以上
話者交替	交互限定, 任意
タスク指向	Yes, No
ドメイン限定	Yes, No
対話履歴の利用	Yes, No
基盤化	Yes, No
意図推定	Yes, No
データベース利用	Yes, No

## 3. FL+PBL 枠組みの実践

この章では, 2.3.3 節で説明した FL+PBL 枠組みの具体的な実践例として, 「発話内の語句が曖昧であった時に人間らしく応答する対話システム」に関するテキストアノテーション演習授業について報告する\*7。

### 3.1 対象の言語現象

演習授業を設計するにあたり, 今回は, テーマとして, 「対話データへのアノテーション」を選択した\*8。

文献 [2] を精読し, 対話システムに関する演習を設計するにあたり, 考慮すべき重要な項目について表 2 のようにまとめた。演習の難易度を考慮しながら, それぞれの項目について表の右側の下線のように設計した。

対象の言語現象としては, 次を選択した。

対話相手の発話内に, 曖昧な語句が含まれている場合, 人間はどのように考えてどのように発話するか

下線部分は, 対話履歴のテキストデータに陽に現れないので, 人手でアノテーションする価値がある情報である。

一般に, テキストアノテーション作業では, 作業者はラベルを付与することが仕事であり, テキストは最初から与えられていることが多い。本演習授業では, アノテーションの対象となる対話テキストについても, 作業者に作成してもらうこととした\*9。

2 章で説明したどの枠組みを使用したかに関わらず, また, 言語学のためであるか自然言語処理のためであるかに関わらず, アノテーション作業は, 一般に, 他人が書いたテキストにラベルを付与する。これは, **読み手によるアノテーション**であり, 読み手の推測により, ラベルを付与する作業である。一方, 自分が書いたテキストに自身でアノテーションすることは, **書き手によるアノテーション**であ

\*7 この章で紹介した演習授業の資料は, 次の URL で一般公開している。

<http://www.cl.inf.uec.ac.jp/lec/>

\*8 このようにテーマを設定した理由は, 2016 年度のオープンキャンパスや研究室公開で, 当研究室を訪問してくれた大学生や高校生が多く, 計算機との対話に興味を持っていたからである。筆者が対話の研究をしているからという理由ではない。

\*9 実際にどのような作業をするのかについては, 3.3 節で説明する。

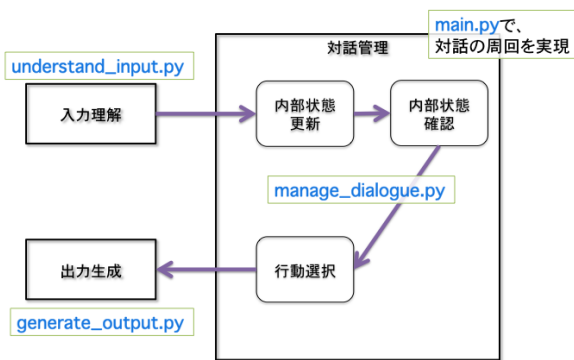


図 1 提供する対話システムの概要

る。読み手によるアノテーションでは、極端な場合を挙げれば、読み手の前提知識や解釈の個人差により、様々な異なるラベルが「正解」として付与されうる。一方、書き手によるアノテーションでは、作業者がラベル体系をしっかり理解しているという前提では、唯一の正解ラベルのみが付与される。ここに誤りは発生しない。

ここ数年、自然言語処理の分野では、自然言語生成の研究が盛んになっている。自動生成の研究を促進するためにも、書き手によるアノテーション済みのテキストデータは構築する価値のあるものであると思われる。

### 3.2 簡易版 NLP システム

文献 [2] における対話システム設計を参考にして、簡単な対話システムを実装した。このシステムの概要を図 1 に示す。最終的に履修者が目指すものは、発話内の語句が曖昧であった時に人間らしく応答する対話システムである。

提供する対話システムは、次の 4 つのモジュールから構成される。

**メイン** 対話のループを管理し、必要に応じて他のモジュールを呼び出す

**入力理解** 入力発話を形態素解析する。曖昧な語句が存在する場合、キーワードとして抽出する

**対話管理** 履修者が改良する

**出力生成** テンプレートを用いて応答を生成する

システムは、次の 2 つの外部辞書を利用できる。

**定型応答辞書** 定型句に対して定型句を応答するための辞書。例えば、「こんにちは」に対して「こんにちは！」を返すように、エントリーを追加できる

**あいまい語辞書** 複数の意味を持つ語に対して、その意味を列挙した辞書。例えば、「ワンピース」に対して、「上着とスカートが一続きになった女性・子供服」や「尾田栄一郎作の漫画」などの意味を返す。エントリーを自由に追加できる

デフォルトのあいまい語辞書は、日本語 Wikipedia\*<sup>10</sup> に存在するすべての曖昧さ回避ページを整形することにより構築した。システムは、入力発話中に曖昧な語句が存在するかどうかをこの辞書を用いて認識する。

デフォルトの出力テンプレートは、以下の 4 つである。

- (1) 定型応答辞書が利用された場合、応答の文字列を出力する
- (2) 曖昧な語句が存在する場合、その意味を縦に並べてすべて列挙する
- (3) 辞書が 2 つとも利用されなかった場合、文頭に近い名詞を 1 つ抽出し、「(名詞) が好きなんですね。」と出力する
- (4) それ以外の場合、「なるほど。」と出力する

履修者には、主に、対話管理モジュールとあいまい語辞書を更新してもらうことになる。定型応答辞書は、プログラミングが得意でない履修者に対する救済措置でもある。

### 3.3 アノテーション例とアノテーション指針

目的の対話システムを開発するために必要となるアノテーション付きテキストデータの例を図 2 に示す。実践では、これともう 1 例の計 2 例を履修者に提示した。

履修者には、この例のように、「曖昧な語句を含むユーザーの発話とそれに対するシステムの応答」を中心とした対話を作るように指示した。履修者への指示は以下のとおりである。

- 「発話」にそのような対話を記述する
- 対話の長さに制限はない
- 曖昧な語句を含む発話に対して、「曖昧な語句」欄にその語句を、「意味 N」欄にそのすべての語義を記述する
- ユーザー発話への適切な応答を生成するにあたり必要となる理想的なシステムの内部動作を、「説明」欄に記述する

「説明」欄に書くべきラベルが、3.1 節で「書き手によるアノテーション」と呼んだ事項である。「説明」欄に履修者が適切なラベルを付与するためにヒントとして、次のように伝えた。

- 人間の頭の中で実行されていると思われる過程を明確にする
- もっともらしい説明で記述することを目指す
- 各人で、異なる過程が動いていることもある

それゆえ、履修者は、「人間の頭の中で実行されている」という観点と「理想的なシステムの内部動作」という観点から、この説明を書くことになる。前者の観点のみからアノテーションする場合、戸惑ってしまう履修者も少なくないと思われるが、後者の観点も示すことにより、見通しがよ

\*<sup>10</sup> <https://dumps.wikimedia.org/jawiki/>

発話者	発話	説明	曖昧な語句	意味 1	意味 2
システム	会話を楽しみましょう。何も入力しないで Enter を押すと、会話を終了できます。				
ユーザー	こんにちは				
システム	こんにちは!				
ユーザー	私はとてもワクワクしています。				
システム	なるほど。				
ユーザー	そういえば、発売したばかりの「ワンピース」を買ってきました。		ワンピース	上着とスカートが一続きになった服。	尾田栄一郎の漫画。主人公はモンキー・D・ルフィ。
システム	発売したね。	どちらか分からないので、無難に返答する。			
ユーザー	スカート部分が長め。	キーワードにより、どちらの意味か分かる。			
システム	私も、上着とスカートが一続きになった服が好きです。	辞書から文字列を取ってきて返答する。			
ユーザー	白色がいいよね。				
システム	白色が好きなんですね。				

図 2 アノテーション例

くなるものと思われる。

### 3.4 最終課題

最終課題として次の 2 種類を課した。

- (1) 曖昧な語句を見つけた時の対話例を 3 つ
- (2) 曖昧な語句を見つけた時に人間らしい応答を返す対話システム

のちに一般公開する可能性を考慮して、(1) に対して次の制約を課した。

- 適切な参考文献を挙げる
  - 公序良俗に反する題材は避ける
  - 個人や団体等を誹謗中傷する題材は避ける
- (1) に対する以下の加点項目を明示的に伝えた。
- 曖昧な語句がちゃんと曖昧である
  - 題材が適切である
  - 応答が冴えている
  - オリジナリティが高い
  - 各発話が自然な日本語である
  - 略称や顔文字などの利用
  - 自然な対話である
  - 参考文献が適切である
  - システムがかわいい、もしくは、かつこい

これらの項目は、最終的に得られるテキストデータが、多様なジャンルを含み、かつ、口語的な表現が多く含まれるように設計した。

(2) に対する以下の加点項目を明示的に伝えた。

- デフォルトのプログラムではできないことができる
- Excel ファイルの対話例が再現できる
- 再現は多ければ多いほど良い

- 対話例にないほうの意味で語句を入力しても対応できる
- 対話例の分析結果から、適切なアルゴリズムが選択され、実装されている
- あいまい語辞書のエン트리記述が適切である
- 定型応答辞書に工夫が見られる

これらの項目は、自身のプログラミング能力に応じて、履修者が 60~100 点の任意の点数を目指せるように設計した。

### 3.5 授業のメタ説明

本章で述べた演習授業は、電気通信大学情報理工学部総合情報学科 3 年生後期の必修専門科目である『経営情報学実験』で実践した。

年度は 2016 年度であり、のべ履修者は 50 名であった。2 グループに分かれて実施する演習であり、先に演習を実施した A グループは 27 名、後に演習を実施した B グループは 23 名であった。

各グループごとに、2 日間のべ 360 分の演習授業を行った。2.3.3 節の「演習段階」で述べたとおり、演習授業は PBL 形式で進め、分からないところを別の履修者と相談することは許可した。アノテーションやプログラミングに関して分からないところを TA に尋ねることも許可した。TA でも対処しきれない細かい問いについては、筆者が対応した。

## 4. 得られたテキストデータ

演習授業終了後、履修者 50 名から、のべ 148 個<sup>\*11</sup>のアノテーション済み対話テキストデータが集まった。そのす

\*11 履修者のうち、2 名は 2 個のみの提出であったため。

べてが、3.3節で望んだ、「曖昧な語句を含むユーザーの発話とそれに対するシステムの応答」を中心とした対話であり、指針設計者が履修者をほぼ管理することなく、目的に適った質の高いテキストデータを構築することができた。

#### 4.1 分析

このデータにおいて、総発話数は、「システム」と「ユーザー」を合わせて、1,691個であり、これらを MeCab<sup>\*12</sup> version 0.996 と UniDic<sup>\*13</sup> version 2.1.2 で形態素解析したところ、16,481形態素であった。このデータには口語的な対話が多かったため、解析誤りが少なからず含まれていると思われる。それゆえ、この値は目安である。

履修者が作例に使用した曖昧な語句は多様性に富んでいた。曖昧な語句の例として、日本語 Wikipedia の曖昧さ回避ページを提示している影響もあり、サブカルチャーに関する語句が多かった。すぐに思いつく曖昧な語句といえば、人名や地名であるが、思いの外、それらを使用した例は少なかった。

実践する前は、「説明」欄のラベルに多様性があることを期待したが、実際にデータを観察すると、事前に提示したラベルを使い回す事例がほとんどであった。それゆえ、ラベルに誤りはなく、質の高いデータになったと言える。今後は、自動集計が可能となるように、書き手によるアノテーションのラベルを選択式にすることを検討してもよいかもしれない。

#### 4.2 利用方法

得られたアノテーション済みテキストデータは、少なくとも以下の4種類の利用方法がある。

1つ目は、純粋に対話データとして利用可能である。タスク指向であっても、非タスク指向の雑談であっても、ユーザーの発話に曖昧な語句が含まれている場合、対話システムは、その問題に対処して発話する必要がある。その機構を実装するにあたり、このデータはとても有用であると思われる。

2つ目は、「説明」欄のラベルを有効に利用する方法である。「発話」欄と「説明」欄の情報を合わせると、対話相手の発話内に曖昧な語句が含まれている場合、人間はどのように考えてどのように発話したかが分かる。本研究では、「説明」欄は書き手アノテーションであるので、「読み手による推測」が入る余地はない。対話システムの評価において、対話システムが出力した発話に対して、あえてメタ質問「どうしてそう言おうと思ったの?」を投げかけることにより、対話システムの人間らしさを評価<sup>\*14</sup>することがで

きるように思われる。「説明」欄の文字列とシステム発話の文字列の類似度を自動的に測定することができれば、上記の人間らしさの評価を自動で実施できると思われる。

3つ目は、得られたデータには口語的な表現が多く含まれていることから、くだけた日本語の解析や生成の研究に利用できる。

4つ目は、このデータは、ストーリー生成の研究にも利用可能であるように思われる。とても短い対話ではあるが、そのほとんどの対話において、曖昧な語句を中心とした寸劇が繰り広げられている。このデータにおける対話の展開を分析することにより、短いストーリーを自動で生成するための知見が得られる可能性がある。

#### 4.3 一般公開のために必要な手続き

今回得られたアノテーション済みテキストデータは、できる限り、一般公開したいと考えている。そのために必要となる手続きは、大きく2つある。

**著作権移譲** 一般に公開し研究教育利用してもよいか、商業利用も可能であるか、著作権者である履修者から許可を得る

**加工** 他人の著作権に関わる文字列などを、適宜、メタタグなどに置換する必要がある。例えば、以下のようなものに適切に対処する必要がある

文章の引用、歌詞の引用、(特に他人を誹謗中傷する)時事ネタの使用、実在する人間の私生活の話題か架空の話か

#### 5. おわりに

本研究では、テキストアノテーションに関する新しいパラダイムを提案した。具体的な実践例として、「発話内の語句が曖昧であった時に人間らしく応答する対話システム」に関するテキストアノテーション演習授業について報告し、得られたアノテーション済みテキストデータが様々な利用方法を持つことを考察した。

今回紹介したアノテーション済みテキストデータを増やす作業は、教育学に詳しくなくとも、3章で説明した授業資料を用いれば、誰にでも可能である<sup>\*15</sup>。提案する枠組みにおいては、アノテーション済みデータの増量に指針設計者に頼る必要はない。一方、3章で説明した演習は、単にアノテーションをするだけであるならば、過剰な技巧である。このデータが有用であり、かつ、ラベル体系やアノテーション指針に問題がないようならば、今後は、アノテーションのみを実施してもよいかもしれない。

演習が完了したばかりの履修者は、このタスクに関してアノテーションするコツを身につけている。優秀な履修者を雇用し、アノテーション作業のみに専念させれば、アノ

<sup>\*12</sup> <http://taku910.github.io/mecab/>

<sup>\*13</sup> <https://ja.osdn.net/projects/unidic/>

<sup>\*14</sup> もちろん、この評価値が高いことは、システムの人間らしさが高いことの必要条件であるが、十分条件ではない。

<sup>\*15</sup> もちろん、履修者、演習環境、TAを用意する必要がある。



セッション済みデータを用意を増やすことができるように思われる。

本枠組みを用いた授業設計の敷居を下げるためにそのノウハウを蓄積することは、今後の課題であり、これからも筆者はこのような演習授業を実践するつもりである。そのノウハウを活かすことができれば、高校や中学の情報授業において、生徒に自然言語処理という分野が存在することを伝え、「自然言語処理は楽しいものである」と早期に興味を持ってもらうことも可能であるかもしれない。

## 謝辞

本研究を遂行するにあたり、電気通信大学情報理工学部総合情報学科の2016年度『経営情報学実験』履修生に感謝の意を表します。この演習授業において、ティーチングアシスタントとして活躍くださいました電気通信大学内海研究室の四十内 佑介さん、今野 慎太郎さん、畠山 幸太さんに深く感謝します。

## 参考文献

- [1] 同志社大学 PBL 推進支援センター：PBL 導入のための手引き, [http://ppsc.doshisha.ac.jp/attach/page/PPSC-PAGE-JA-9/56858/file/pblguidebook\\_2011.pdf](http://ppsc.doshisha.ac.jp/attach/page/PPSC-PAGE-JA-9/56858/file/pblguidebook_2011.pdf) (2011).
- [2] 中野幹生, 駒谷和範, 船越孝太郎, 中野有紀子：自然言語処理シリーズ7 対話システム, コロナ社 (2015).
- [3] Nowak, S. and Rüger, S.: How Reliable Are Annotations via Crowdsourcing: A Study About Inter-annotator Agreement for Multi-label Image Annotation, *Proceedings of the International Conference on Multimedia Information Retrieval*, ACM, pp. 557–566 (2010).
- [4] Pustejovsky, J. and Stubbs, A.: *Natural Language Annotation for Machine Learning*, O'Reilly Media (2012).
- [5] Sorokin, A. and Forsyth, D.: Utility data annotation with Amazon Mechanical Turk, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2008*, pp. 1–8 (2008).
- [6] 武田俊之：反転授業に関する実践および研究の展望, 関西学院大学高等教育研究第6号, pp. 95–100 (2016).
- [7] von Ahn, L. and Dabbish, L.: Labeling Images with a Computer Game, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 319–326 (2004).