

自然言語処理の未解決課題 ～対話、法律、医療、政治分野での 研究事例

第7回特許情報シンポジウム 招待講演

狩 野 芳 伸
Yoshinobu Kano
静岡大学 情報学部
行 動 情 報 学 科



国立大学法人
静岡大学

Kano Laboratory

National University Corporation
Shizuoka University

自己紹介: 狩野 芳伸

- 静岡大学 情報学部 行動情報学科 准教授(2014年～)
 - 浜松キャンパス
- 修士課程以来「自然言語処理」が専門
 - 学部: 東京大学理学部物理学科
 - 修士～博士～ポスドク: 情報理工学系研究科コンピュータ科学専攻(辻井潤一先生)
 - JSTさきがけ研究者「情報環境と人」領域
- 人间的に言葉を「理解・生成・対話」するシステムの構築
 - 音声・テキストを統合したリアルタイム解析器
 - 機械学習一発ではできない、過程を説明可能なホワイトボックス
 - 仕組みも人間に近づける: 人間(と脳)自体の仕組みを考察
 - 言語だけでなく背景となるすべての知的処理を包含する
 - 箱庭ではなく実応用で評価: 対話システム、質問応答、文生成、法律(試験解答)、医療(診断支援)、政治など
 - ※特許を扱ったこともあるのですが、今日はあえてそれ以外を
- 詳細は <http://kanolab.net/kano/index.ja.html>

進行中の主な研究テーマ・プロジェクト

- 心理学的に妥当な言語処理モデルと応用 (JSTさきがけ、萌芽)
 - 脳科学・認知科学による人間に近いモデルに基づく日本語話し言葉解析器の構築と検証 (挑戦的開拓 東京大・上智大)
- 対話システム: 人狼知能プロジェクト (東京大 鳥海先生ほか) 自然言語部門オーガナイザ
- 意味の重ね合わせによる高度な言語生成 (吉田英雄記念事業財団)
- 法律文書の処理と裁判の自動化支援
 - 試験問題の自動解答・質問応答 (東ロボ: 大学入試の自動解答): 終了
 - COLIEE 司法試験自動解答の国際タスク (基盤S NII 佐藤健先生)
- SNS・新聞記事・議会議事録を用いたAIによる世論形成過程と政治家の応答性の分析 (基盤B 学習院大・慶応大・大阪大)
 - 議会議事録の自動分析 (挑戦的開拓 学習院大 野中先生)
 - SNSにおける欺瞞とその広がり自動検出・推測と政治学・社会学的分析および予防的介入 (セコム科学技術財団)
- 医療言語処理: 電子カルテの処理と自動診断
 - 精神疾患の自動診断 (CREST/AIP加速 共同研究G 慶応義塾大学病院岸本先生/NII 佐藤真一先生)
 - 効果・副作用の発見と治療方針 (浜松医科大・国立病院機構)
 - 放射線学会・病理学会の収集する全国所見データ (AMED JP-AID)

言語処理に残された課題は何か

- 超大規模深層学習モデルによる性能向上
 - 特に近年の滑らかな文生成は目を見張るものがある
 - しかし、訓練データもパラメータ数も膨大
 - 人間には一生かけても不可能な学習
 - 人に近い振る舞い・結果を得たいならば、仕組み(制約)も人間に近づけるべき
 - 自然言語は人間の制約の下で形づくられているはず
- 課題: 計算資源と実行時間のコスト
 - トータルで人間よりはるかに計算量を要する
- 課題: 計算過程の解釈が難しい
 - 応用によっては説明や根拠が必須
- 課題: 首尾一貫した生成ができていない
 - 当然、論理性も疑わしい
- 課題: 「常識」「世界の知識」が扱えていない
- **いまの技術で解きうる課題**にも取り組みつつ、**解けていない課題に挑戦**してきた
 - その概要をご紹介します: 医療・対話・法律・政治
 - 分野の今後をご一緒に考えたいと思います

電子カルテを利用した医療支援

- 大規模機械学習によるがん治療の最適化
 - 抗がん剤の作用副作用、患者属性を用いた腫瘍マーカー値の時系列推測
 - 国立病院機構との共同研究、数百万件規模の電子カルテを集積した大規模医療集積基盤(NCDA)を利用
 - 実は絞るとデータ量が足りない
 - 数値データのほうが有効な可能性
 - その中の着目ポイントを教えてくれる？(医師の知見)
 - 電子カルテテキストは相当broken
 - 数値・テキストの解釈に各科の専門知識が必要
- 「カルテ」からの「退院サマリ」の自動生成(厚生労働省科研費)
 - 要約のようで要約でない
 - カルテに書いていないことも書かれる
 - 医師・病院により大きく異なる
 - 「よいサマリとは何か」を指導する勉強会が院内で開かれる…

自然言語処理による精神疾患自動分類

- 話し言葉
 - 会話の録音から精神疾患を予測する
 - 現状、1000時間程度の世界最大規模診断付き音声コーパス
 - 1名1visitあたり1時間程度の録音
 - 1時間は3パート(質疑自由会話・絵の説明・物語の説明)に分かれる
 - 全データ文字起こし済み
 - 一部データに詳細な音声・言語学的アノテーションを付与
- 書き言葉
 - SNS投稿の書き込みから精神疾患を予測する
- 研究プロジェクト
 - JST CREST「自然言語処理による心の病の理解:未病で精神疾患を防ぐ」(共同研究G)
 - JST CREST「精神医学×メディア解析技術による心の病の定量化・早期発見と社会サービスの創出」(加速フェーズ)(共同研究G)
 - AIP加速「精神医学×メディア解析技術の展開:精神疾患への介入の挑戦」(共同研究G)

自然言語処理による精神疾患自動分類

- 精神科の診断は問診が基本のためまさに言語処理～会話分析だが、自然言語処理としては難しい課題
 - うつ病: 悲観的発言が増え、思考が遅くなり、質問のゴールにたどり着きづらくなる(精神運動制止)
 - 躁病: 注意を集中させることができず、考えが飛躍する(観念奔逸)
 - 統合失調症: 猜疑的な発言が増え、文脈や単語のつながりが弱くなる(連合弛緩)
 - 認知症: 物の名前が出にくくなり(語想起障害)、指示語が増え、話が回りくどくなる(迂遠)
 - 一方で診断基準は必ずしも安定しない
 - できれば「未病」に対応したい
- 現在のところ、各疾患vs健常の分類でそれぞれ8～9割程度の性能
 - 発達障害(ASD)のスコアリングでも高い性能を達成

健常者との比較観察

- 共通: 長い文は難しい傾向、文が複雑になるから？
 - 接続詞が少ない、名詞の語彙
- うつ病: 「自己主張」が強い？
 - 主格の“ハ格”が多く見られ、語彙も多い
- 認知症: 対話能力が落ちている
 - 間違いが多く、フィラーとしての指示語も多いが、内容語は少なく、応答も非内容語が多い
 - 発話音量平均が大きい、おそらく加齢による聴覚の低下による
- 統合失調症: 物語説明パートで、「証拠」に乏しい？
 - 目的格である“二格”が少なく、補語も少ない

書き言葉（SNS投稿）の分析

- Twitter等のSNS投稿を分析
 - メンタルヘルスに不調のあるユーザを推定したい
- 超大規模データを収集
 - 半自動ラベル付与
 - 時系列変化を追う
 - ユーザ間の影響、社会的なイベントの影響を分析
- SNS投稿テキストの分析は難しい
 - 短文が中心
 - 省略がおおい
 - 文脈依存性が高い
- 一方で大規模データが取得可能
 - タスク設定の工夫次第で興味深い研究ができそう

アカウントを不調とみなすパターン

- 精神疾患に使用される薬剤名を含むツイート履歴がある
 - 79の抗精神病薬、抗うつ薬、抗不安薬（右上表）
- メンタルヘルスを示唆する単語（右下表）をプロフィールに含む
 - 薬剤名および示唆する表現を含むツイートは除外
- この2条件に合致するアカウントを抽出し、不調（正例）とみなす
 - アカウント数：15,266人
 - ツイート数：34,505,381件
- 非不調群もほぼ同数に揃えた
 - アカウント数：16,519人
 - ツイート数：28,685,425件
- 正例負例ともノイズが入りうるが、大規模なため統計的に意味のある処理ができると期待

表3：使用した抗精神病薬、抗うつ薬、抗不安薬の薬剤名（五十音順）

アナフラニール、アピリット、アモキシサン、イフェクサー、イフェクサーSR、インプロメン、インヴェガ、ウインタミン、エビリファイ、エミレース、エリスパン、クレミン、クロフェクトン、クロルジアゼポキシド、コレミナール、コンスタン、コントミン、コントロール、サインバルタ、ジアパックス、ジェイゾロフト、ジプレキサ、スピロピタン、セディーエル、セトウス、セニラン、セパゾン、セルシン、セレネース、セロクエル、ゼプリオン、ソラナックス、ゾテピン、タンドスピロン、ダイアップ、チミペロン、テトラミド、デジレル、デパス、デプロメール、トフラニール、トリプタノール、トレドミン、トロペロン、ドグマチール、ニューレプチル、ネオベリドール、ノバミン、ノリレン、ハロマンズ、バルネチール、パキシル、ヒルナミン、ピブレzzo、ビーゼットシー、フルデカシン、フルメジン、プロピタン、ホリゾン、ホーリット、ミラドール、メイラックス、メレックス、メンドン、リスパダール、リフレックス、ルジオミール、ルボックス、ルーラン、レキサルティ、レキソタン、レクサプロ、レスタス、レスリン、レボトミン、レメロン、ロドピン、ロナセン、ワイパックス

表4：気分（感情）障害圏、統合失調症圏及び神経症性障害、ストレス関連障がい及び身体表現性障害圏の訴えや症状に関連する単語

うつ、鬱、躁、パニック、不安、ネガティブ、精神病、精神の病、自傷、障害者手帳、倦怠、動悸、息苦、食欲不振、頭痛、吐き気、しびれ、痺れ、麻痺、過眠、食欲低下、散漫、めまい、目眩、息切れ

メンタルヘルス不調群のツイート例

- 顕著なものとうそでないものがあり、人間には明確にわからないものも検出できる可能性がある
 - 実験では250トークンor120トークンに連結して使用

メンタルヘルス不調群のツイート例

※文脈情報を失わない程度に手作業で匿名化(改変)した

顕著	死にたくても死ねない。もう何もわからない。いつも親と仕事の話してるんだけど、無意識に死にたいとか言って親泣かせてる自分がうざいし、死んだほうがまし。で、死にたいとか考えると逆に落ち着いてくる。けど生きててもなあ、っていうループにはまって結局死にたいってなる。
曖昧	風呂に入りながら娘の帰宅を待つが、帰ってくる気配なし。手元にあるスマホで聞いてみようか。ご飯も出来てるしはよ帰ってきてほしい 何してんだろう眠くなってきた。でも外のバイクうるさすぎるから寝落ちすることはない、まあリラックスタイムを満喫しますか
隠微	おはようございます！寒すぎて散歩に行くか迷ってるけど、犬つれてけば暖かいかな、、でもちょっと貧血気味だからやめとこうかな汗 でも犬は歩かないわけにはいかないって顔してるし、ちょっと出かけます！

SNSにおける欺瞞とその広がり of 自動検出・推測と政治学・社会学的分析および予防的介入

- **超スマート社会の「悪」を探るには必須：SNS**
 - 情報化社会において、テキストメッセージは作成・伝達のコストが低く、人間への影響力は大きい e.g. フェイクニュース
- **欺瞞とその広がりを自動検出し予防的介入を試みる**
 - 欺瞞：他者を意図的に誤った方向に導く
 - － 「嘘」は「欺瞞」の一種で間違っていることが分かっている情報を言って他者をだますこと
 - － 欺瞞を意図的でなく伝達してしまうことも「悪」に加担する
 - 「虚偽」かどうかは、「事実」に反するかどうかで決まるが、何が事実かの機械的判定は困難なので、**相対的な関係**に留める
 - － 同様の情報の「流れ」～「流れ」同士の矛盾～どちらかが欺瞞
- **マクロな情報の流れと変化をミクロな言語表現から組上げ可視化・推測する**
 - 既存研究は**単語統計**しか見していない
 - 基盤となる道具：文章ペアの**関係認識**(矛盾・含意・中立)と影響力予測を**ミクロに計算、全体を時系列でとらえる**

SNSにおける欺瞞とその広がりの自動検出・推測と政治学・社会学的分析および予防的介入(セコム科学技術財団研究助成 代表)

(悪の) インフルエンサー
どういう人が
どう発信すると
誰にどのように届くか？

(欺瞞に) 流される大衆
どういう人が
どう受信すると
どのように変わるか？

欺瞞の発信

欺瞞の受容・伝達

欺瞞の防止

悪

自己矛盾した投稿

「事実」と矛盾した投稿

ユーザ
ネットワーク
フォロワー
フォロワー

一般ユーザ

リツイート

いいね

変容した内容
の情報伝達

訂正
発信
実験

欺瞞の伝達防止と 訂正効果分析

影響推測器を用いた
影響力の大きい
文面生成

意見変化につながる
発信者・受信者
属性と文面の分析

ツイート

ニュース記事

国会議事録

捉えたい欺瞞情報の流れ

入社禁止期ソノウドソーシング調査

心理学的属性 政治/社会的意識
学歴 職業 年齢 家庭環境 等

アノテーション付与・ファインチューニング

欺瞞と広がりの検出器

話し言葉省略補完・解析器

気分・心理状態推測器 含意関係認識器
嘘検出器 影響（リツイート）推測器

事前学習

チューニング

大規模
ツイート
データ

科研費「～AIによる世
論形成過程と政治家
の応答性の分析」
関連成果

政治学・社会学的分析

イデオロギー測定と政治的傾向
移民の参政権・生活保護問題の分析 等

アカウント 時系列変化

時間

受信
発信
(投稿)
アンケート
調査
...

基盤的言語処理ツールの構築

- 本研究の入力はSNS投稿・ニュース記事・国会議事録
- **SNS投稿の主語・目的語補完システムを構築済み**
 - SNS投稿は**省略・崩れ**が多く話し言葉に近い
 - 大規模ツイートデータによる事前学習時に省略補完のバイアスをかけることで高精度で省略補完を実現
 - 述語項構造(文内照応・文外照応・談話外照応・係り受け)で80%を超える性能:**本研究で省略補完に利用する**
 - 応用として対話補完性能、対話システム性能(一貫性)の主観評価において既存システムを大きく上回った(発表予定)
 - 構築した事前学習済みモデルは汎用のため、**本研究のすべての言語処理タスクに利用し照応の性能向上を図る**
- ほか、下記研究プロジェクトで構築している事前学習済みモデルも統合する

心理学的に妥当な言語処理モデルと応用(JSTさきがけ、科研萌芽 代表)

脳科学・認知科学による人間に近いモデルに基づく日本語話し言葉解析器の構築と
検証(科研挑戦的開拓 東京大・上智大 代表)



嘘つきを見破る：人狼知能

- 人狼知能プロジェクト(後述)
 - 人狼ゲーム：嘘つき役「人狼」とそれ以外「村人」に分かれ、会話を通じて隠された役を見抜く
 - 人狼を自動プレイさせるプロジェクトを数年にわたり主催中
- 人狼ゲームの掲示板大規模コーパスを用い、**嘘つき(人狼)の自動検出**を行った
 - 7～8割程度の予測性能
 - 内容や論理以外の部分(広い意味での文体や会話パターン)で推測
- **精神・心理状態推測器(前項)と合わせ「人物タイプ」を推測**
 - **後述のビッグファイブ・ダークトライアドの推測も合わせて学習**
 - ビッグファイブ：**パーソナリティ特性**(開放性、誠実性、外向性、協調性、神経症傾向)
 - ダークトライアド：**反社会性パーソナリティ**(自己愛症：ナルシシズム、権謀術数主義：マキャヴェリアニズム、精神病質：サイコパシー)



世論形成過程の推測

■ 世論形成過程と政治家の応答性の分析

- 過去の全国会データを用いた国会議事録の分析
- 政治家の発言の含意関係認識(含意:相手文の意味を包含するか)
 - 含意関係認識器の絶対的性能が不足～大学入試社会科自動解答、司法試験自動解答・裁判の自動化支援
- 「バズリ」リツイート数～影響力の回帰学習による予測
 - 政策的世論へのインフルエンサーを抽出、フォロワー・フォロワー数で正規化してから予測を試みる
- Twitterユーザのクラウドソーシングによる大規模調査
 - パーソナリティ、心理学的属性、政治的傾向、基本個人属性など
- 政治学・社会学的分析と予防的介入実験
 - イデオロギー推測、対外国人意識・移民問題と虚偽情報など

SNS・新聞記事・議会議事録を用いたAIによる世論形成過程と政治家の応答性の分析
(基盤B 学習院大・慶応大・大阪大)(代表)

試験問題の自動解答・質問応答(東ロボ:大学入試の自動解答)(社会科担当)COLIEE
司法試験自動解答(基盤S分担 国立情報学研究所)(U Alberta)主催者

欺瞞の発信

悪

自己矛盾した投稿

「事実」と矛盾した投稿

ツイート

ニュース記事

国会議事録

欺瞞の受容・伝達

ユーザ
ネットワーク
フォロワー
フォロワー

一般ユーザ

リツイート

いいね

変容した内容
の情報伝達

訂正
発信
実験

大規模定期クラウドソーシング調査
心理学的属性 政治/社会的意識
学歴 職業 年齢 家庭環境 等

アノテーション付与・ファインチューニング

欺瞞と広がりの検出器

話し言葉省略補完・解析器

気分・心理状態推測器 含意関係認識器
嘘検出器 影響（リツイート）推測器

政治学・社会学的分析

イデオロギー測定と政治的傾向
移民の参政権・生活保護問題の分析 等

欺瞞の防止

欺瞞の伝達防止と
訂正効果分析

影響推測器を用いた
影響力の大きい
文面生成

意見変化につながる
発信者・受信者
属性と文面の分析

アカウント
時系列変化

時間

受信

発信
(投稿)

...

アン
ケート
調査

...

事前学習

チューニング

科研費「～AIによる世
論形成過程と政治家
の応答性の分析」
関連成果

難しい課題

COLIEE: 法律文書の自動処理タスク

- COLIEE (Competition of Legal Information Extraction/Entailment)
 - 具体的なタスクの共有による法律文書処理研究コミュニティ形成が目標: 毎年開催 (COLIEE 2023開催中!)
 - 大きく情報抽出と含意関係認識の2タスクに分割
 - Case Law (判例法、カナダ): Task1, Task2 (COLIEE 2018より新設)
 - カナダ連邦裁判所の判例DBを利用
 - vLex社の提供による
 - Statute Law (成文法、日本): Task3, Task4
 - 司法試験の民法短答式を利用
 - 問題文と法律文の含意関係を二値で答える (Task 4)
 - オーガナイザーは試験問題と法律条文を日英両言語で提供
 - 現在、約1000ペア
 - 毎年最新の問題をテストデータにし過去問は訓練データに追加

COLIEE: 実際の短答式問題の例

- 多択式の問題を、各選択肢ごとにYes/Noの二択に変換
 - 英訳も作成
 - 下記は、「簡単な」問題例

抵当不動産を買い受けた第三者が、抵当権者の請求に応じてその抵当権者にその代価を弁済したときは、抵当権は、その第三者のために消滅する。

(平成26年民法短答式第14問選択肢工)

民法 第三百七十八条

抵当不動産について所有権又は地上権を買い受けた第三者が、抵当権者の請求に応じてその抵当権者にその代価を弁済したときは、抵当権は、その第三者のために消滅する。

COLIEE Task 4 の結果

- BERT系の深層言語モデルとそのアンサンブルを使った手法が上位

Team	L	#	Accuracy	Team	L	#	Accuracy
JNLP.BERTLaw	E	81	0.7232	KIS_3	J	61	0.5446
TRC3mt	E	70	0.6250	sim_neg.OvGU	E	61	0.5446
TRC3t5	E	70	0.6250	UEC1	J	61	0.5446
UA_attention_final	?	70	0.6250	taxi_BERTXGB	E	60	0.5357
UA_roberta_final	?	70	0.6250	UECplus	J	60	0.5357
KIS_2	J	69	0.6161	CUGIVEN	E	58	0.5179
llntu	J	69	0.6161	CUPLUS	E	58	0.5179
cyber	E	69	0.6161	linearsvm_no_ngram.HONto	E	57	0.5089
UA_structure	?	68	0.6071	POS_simneg.OvGU	E	57	0.5089
GK_NLP	?	63	0.5625	taxi_le_bigru	E	57	0.5089
linearsvm.HONto	E	63	0.5625	TRC3A	E	56	0.5000
JNLP.BERT	E	63	0.5625	UEC2	J	55	0.4911
KIS	J	63	0.5625	baseline_attention.OvGU	E	54	0.4821
linearsvm_no_ngram.HONto	E	62	0.5536	AUT99-BERT-MatchPyramid	E	52	0.4643
JNLP.TfidfBERT	E	62	0.5536	AUT99-LSTM-CNN-Attention	E	50	0.4464

開催を通じた分析～主な課題点

- 法律文書特有の構造
 - ～COLIEE 2019: 古典的な言語処理手法
- 絶対的なデータ量の不足
 - ～COLIEE 2020: 事前学習(転移学習)による深層言語モデル
 - ～COLIEE 2022: アンサンブル(システム結果の混合)
- 一般的な知識源(常識、世界の知識)の不足
 - ???

問題解答要素技術ごと統計(上位抜粋)

task	# prob	11	23	24	25	26	5	14	16	27	6	7	10	13	8	12	15
正解数		81	70	70	70	70	69	69	69	68	63	63	63	63	62	62	61
言語		E	E	E	E	E	E	J	J			E	E	J	E	E	J
条件	74	47	46	43	43	39	46	44	45	43	45	42	39	44	40	44	42
人物役割	48	35	29	25	30	30	29	29	29	26	22	25	25	24	24	26	24
人物関係	48	35	29	26	30	29	29	30	29	25	23	25	26	23	24	28	23
形態素解析	33	30	20	22	22	23	21	27	21	24	23	21	26	22	22	19	22
照応解析	23	17	15	15	14	16	12	15	14	13	13	11	14	13	12	10	14
曖昧性解消	1	1	0	0	0	1	1	0	1	1	1	1	0	1	1	0	1
意味役割付与	14	9	9	11	9	9	9	9	10	8	6	5	8	8	5	9	7
動詞言換	41	32	25	26	23	20	25	26	27	22	20	20	24	21	19	21	20
一般辞書	16	12	7	9	10	10	11	9	9	12	9	7	11	9	8	9	9
述語項構造	73	57	45	47	45	48	47	45	46	44	38	39	41	40	39	37	38
否定形	69	47	46	40	45	41	48	40	41	42	35	40	34	42	37	37	40
法律事実関係	55	37	30	34	34	34	33	34	33	32	29	30	29	28	29	32	26
含意関係	20	16	11	11	12	11	9	11	12	11	12	12	7	12	12	11	12
係り受け	24	18	17	16	14	15	16	17	16	16	15	11	15	14	13	12	12
条文検索	12	7	6	9	7	6	6	9	9	5	8	3	9	3	3	10	4
一般言換	3	3	3	2	2	1	2	1	2	3	0	2	1	2	2	2	2
箇条書き	11	6	6	7	5	4	5	7	7	7	4	5	7	5	5	7	5
数値計算	1	0	1	0	1	1	1	0	0	1	0	0	0	1	0	0	24

今後のCOLIEEタスク設計

- 説明可能なAI
 - 参加者は短期間で実装させるため、安易な方向に走りがち
 - End-to-end で学習しておわり が多い
 - 結果が良かったとしても、実用につながるのか？ 本当にちゃんと解けているのか？
 - 根拠を示させるなど、「説明」の要素が入ったタスク設計をしたい
 - 正解の作成、自動評価基準をどうするか？ 説明箇所をあてさせる？
- COLIEEの目標を達成することは、**法律分野のみならず人間の知的能力を幅広くカバーする、人工知能分野の究極的な目標を包含する**
 - 参加者募集中です！

人狼知能：嘘・説得・信頼

■ 人狼のゲーム性

- 客観視点での情報不確定性(不完全情報性)
 - － 得られた情報が真かどうか不明
- 推理：他者の意図のモデル化
 - － 自分がこう思っていると相手が思っているだろう
 - － 文脈からの真偽の判定
- 説得：他者から見た自己のモデル化
 - － 相手に情報を与えて信頼を得る

■ 嘘と信頼

- 誰かが嘘をついていうという状況
- 村人側：自分が語る事をどうやって信用してもらうか
- 人狼側：自分が騙る事をどうやって信用してもらうか
 - － 襲撃する対象をどう話し合って決めるか
- 「説得する人工知能」



人間らしい対話システムの評価として

- GPTベースの大規模言語モデルにより、対話システムは近年急速に性能を上げている
 - 語彙・文法はとても流暢で、1文では人間と見分けがつかないレベルになってきた
 - 一問一答での受け答えも、相当かみ合っている
- まだまだ完璧ではない: 前提知識の共有や深い意味などが取れているかは疑わしい
 - 内容や設定に対する一貫性のなさ
 - 矛盾した応答をするということは、論理性も疑わしい
 - 長い文脈はおそらくあまりとれていない
 - 「常識」「共通基盤」といった外部知識に対応できていない
 - 大規模データの統計的な「重ね合わせ」でどこまでいけるか？
- 人狼なら相手の話を聞かねばゲームにならないはず
 - より人間に近づくほどそうなる(あるいは人間と対戦)
 - 評価は勝敗よりも、対話～ゲーム行動のリンク
- 毎年「人狼知能大会」を開催しているが、まだまだ「会話を通じたゲームがきちんとできている」とは言い難いレベル

「力業」の限界と今後

- 一般的な知識の不足：人間は自然と身に着ける
 - 法律文書に現れるのは最終的な表現であって、背後にさまざまな知識、人間特有の知的構造がある
 - 機械は何兆ものデータを利用しているが人間は小規模でできる
 - 規模もモデル構造もどんどん人間から離れていっている
- 「古典的な」言語処理と深層学習の融合を図るときか
 - 自然言語処理分野全体としても、次なるブレークスルーではないか
 - 「論理構造」「述語項構造」など
 - 深層学習は埋め込みベクトルを使っているが、もっと「決定的」な要素を含める など
 - シンボルと分散表現の融合：e.g. 多義語はどうなっている？
- 人間に近いモデルの導入
 - 逐次的・音声処理の融合、人間に近い制約を加えたモデル 等