DIRECT
Data-driven Intelligent System Research Center

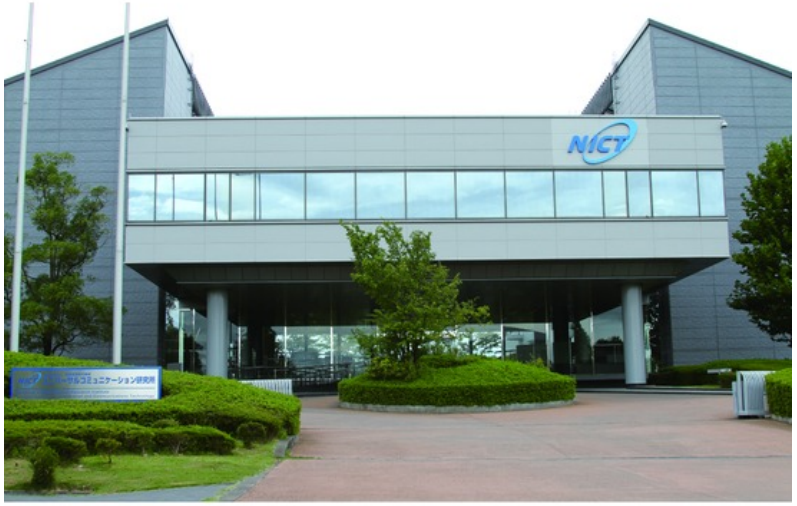
データ駆動知識処理研究室

鳥澤 健太郎 客員教授
飯田 龍 客員准教授

NICT

実体は？？

- 国立研究開発法人　情報通信研究機構(NICT)
  - データ駆動知能システム研究センター
  - NAISTからバスで２５分、車で１０分
    - なにかあれば、すぐ来れる、行ける距離



- 博士号保有のプロの研究員が９名、プログラマ、データ作業補助等１４名、データ作成作業者６０名以上
- 客員教授、准教授は元国立大学教員
- １０００台近い専用サーバー、１０００枚近い専用GPGPU
- Web３００億ページを収集済み、日本語Tweetの10%を常時取得

研究テーマ、研究室の特色

- Web数百億ページ等、大量のデータを使った自然言語処理、特に対話、質問応答、文脈解析技術全般
  - 他の組織にはない大量の学習、評価用データを自由に使って深層学習等の研究をすぐにはじめられる
  - 数十人の作業者に頼んで独自の学習・評価用データを作ることも可能
  - すでに稼働している対話システム、質問応答システム等を研究で自由に利用することも可能

  - システムの実装、データ作成は、プログラマ、言語学者やデータ作成の経験が長い技術者がサポート

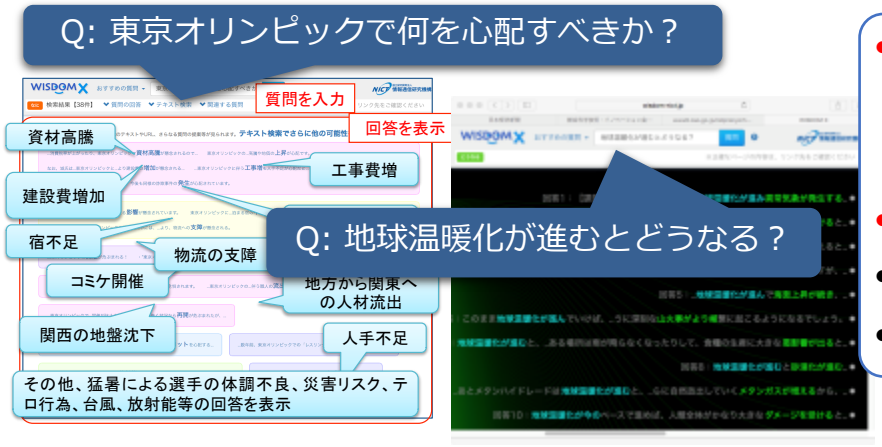
これまでの成果

３つの大規模AIシステムを試験公開中

対災害SNS情報分析システム　*DISAANA / D-SUMM*

- Twitter上の災害情報をリアルタイムに分析し、被害状況の要約や質問が可能
- 九州北部豪雨等で実活用（大分県庁は鉄橋流失を鉄道会社よりも早くD-SUMMで発見し、鉄道会社に通知）
- 民間企業へライセンスも実施。研究員には給料とは別にライセンス料も
- 昨年度はNHKスペシャル等、報道100件以上

大規模Web情報分析システム　WISDOMX



- ４０億件以上のWebページをもとに、「東京　オリンピックで何を心配すべきか?」といった多様な質問に回答や仮説を提示
- 科学論文の内容を先取りする仮説が出た実例も。
- 入力キーワードから、回答可能な質問の列挙も 可能
- 民間企業へのライセンスも実施

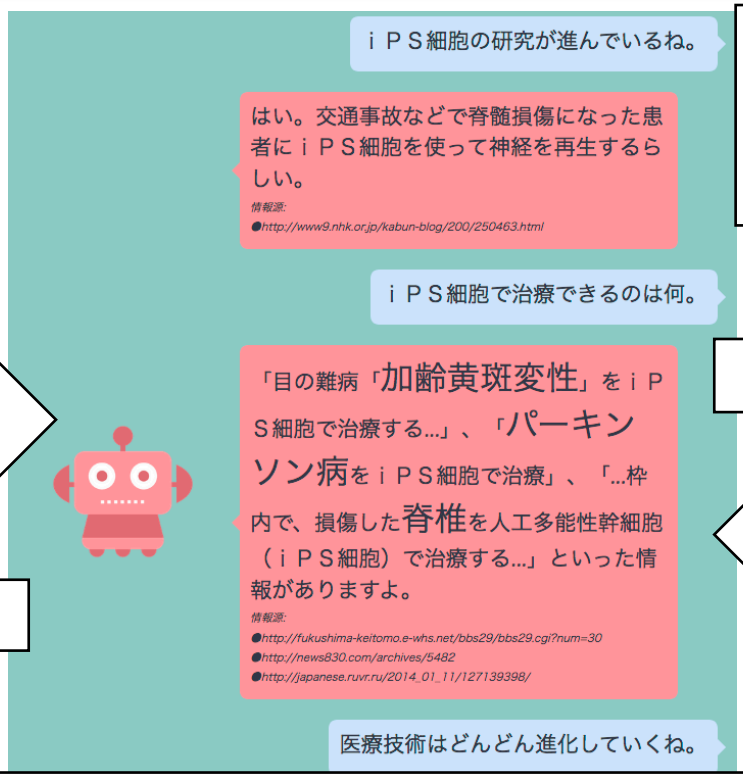
現在研究中

次世代音声対話システムWEKDA

ユーザ

音声認識

音声合成



この質問の合成法でWEKDAの人格、動機、用途が決まる設計（今後の重要な研究課題）

質問を合成

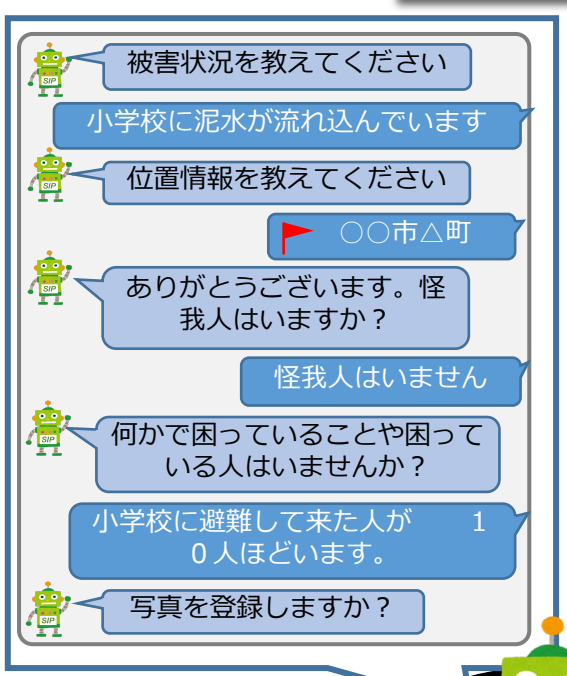
回答

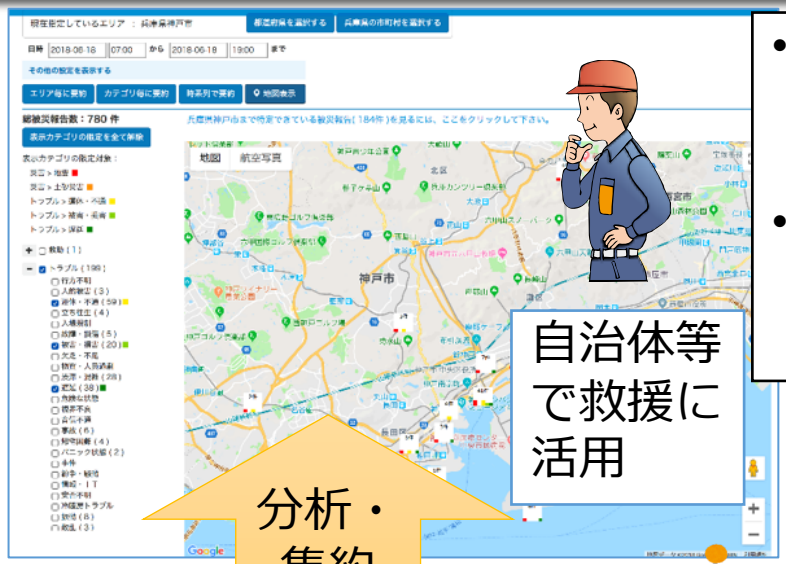
WISDOMX

- Web４０億ページに書かれた知識を使って対話をする博学対話システム
- 基本、プログラマが作り込んだ既存対話システムが「理解できません」と言ってしまうような多様な入力にも応答
- 深層学習の塊、質問の合成法をかえることで様々な用途に適用可能
- いずれは高齢者介護、教育、研究者等のブレインストーミングで活用を狙う

現在研究中

LINE上の防災チャットボットSOCDA

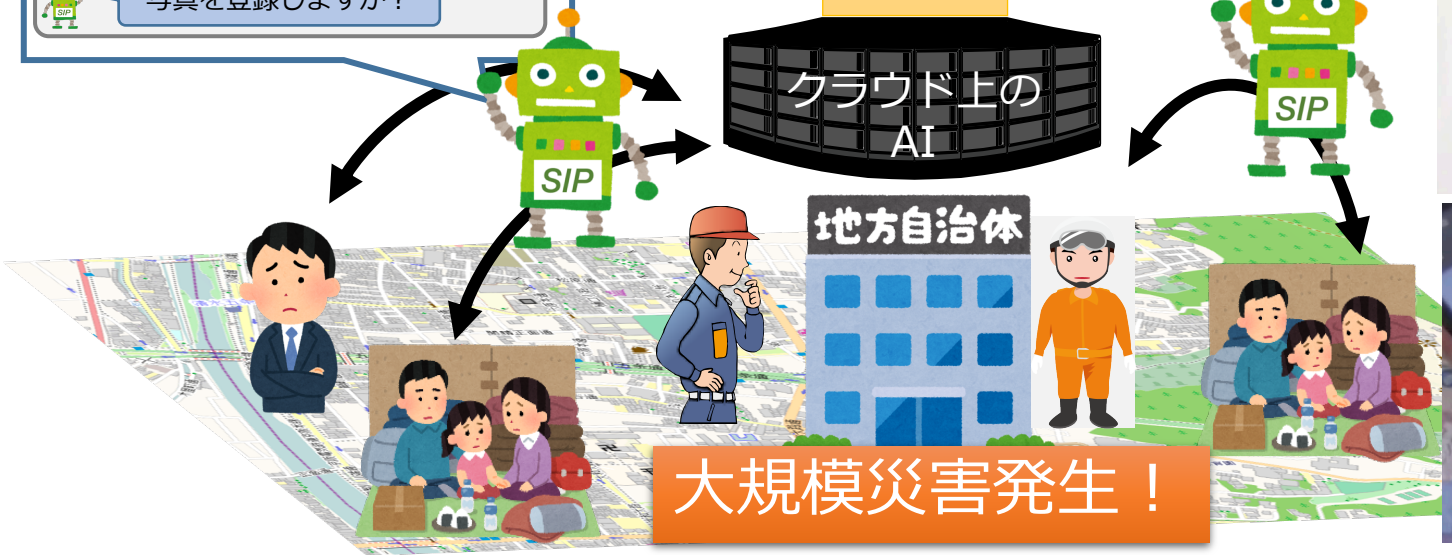




自治体等で救援に活用

- 被災者一人一人にチャットボットが寄り添って被害情報を収集、情報提供
- LINE株式会社、防災科研、ウェザーニューズと連携して開発

NHKスペシャル 1月17日



取り組んでもらう研究テーマの具体例

- 音声対話、質問応答の高度化
- 例：対話システムによりストーリーに沿った雑談を行わせる
  - 子供の学習に役立つ対話のストーリー、高齢者の健康維持に役立つストーリー、研究のブレインストーミングで役立つストーリー
  - ストーリーとは何か？　人が自然に感じるストーリーはどうやって自動生成できるか？
- 例：対話システムにそれまでの対話履歴やユーザの性質を踏まえた対話をさせる
  - 入力中の省略を自動的に推測する
  - 高齢者、子供、アニメおたくと、それぞれ異なる対話を行う
- 例：Webページ等のテキストと深層学習を使い、難しい質問に答えさせる
  - Web等には直接書いていない回答も複数の文書の情報を組み合わせ、仮説を作って回答
  - 現在でも「カワセミの写真はどうしたらうまく撮れるか？」「人工知能が進化を遂げたのはなぜ？」等の複雑な質問にそれなりに回答できるが、より複雑な質問に答えたい
- もちろんこれらに必要となる基礎研究に大規模データを用いて取り組むことも可能
- ビッグデータ、音声対話、深層学習の組み合わせは就職上も有利

生活

- 数学やプログラミングの知識があれば有用ですが、知識がなくても一から勉強できます。教員はそうした人の指導の経験あり
- 最初は、知能コミュニケーション研究室（中村研）で勉強。勉強会への参加も可
- M1後半から、NICTに入り浸り、NICTの設備でどっぷり研究するのもあり。(もちろん、その前からも可)
  - NICTの「協力研究員」に任命。NAIST-NICTのバス代は支給。自動車での行き来も可
  - ACL等にたくさん論文を書いている研究員といつでも議論可能、希望すれば英語でのコミュニケーションも可。PC貸与。本、論文はほぼすきなだけ購入可能
  - 企業との共同研究等への参加や自治体等との連携の現場にでるのも可
- コアタイム等は、授業や生活に合わせて臨機応変に調整
- 博士課程では、NICTの「有期技術員」として給料をもらいながら博士号を狙うことも可能。もちろん、その後、「研究員」に昇格もあり

論文＆受賞

- もちろん、システムを作るだけではない
- ２００８年にNICTで研究室（センター）発足以降
  - ACL 10本、 EMNLP 13本、 AAAI 5本等、 30本以上のトップカンファレンス論文
  - 多数の主要な学術賞を受賞：日本学術振興会賞、日本オープンイノベーション大賞 総務大臣賞、文部科学大臣表彰科学技術賞、文部科学大臣表彰若手科学者賞、ドコモモバイルサイエンス賞、前島密賞、Twitter Data Grants、言語処理学会２０周年記念論文賞、COLING/ACL 2006 Asian Federation of Natural Language Processing Best Asian NLP Paper Award、情報処理学会論文賞、山下記念賞、各学会論文賞等

興味がある方は…

- 「NAIST いつでも見学会」を通じてご連絡ください

- 詳細はこちらから

  https://isw3.naist.jp/Contents/Admission/CampusTour-ja.html