

りんな：女子高生人工知能

Xianchao Wu, Kazushige Ito, Katsuya Iida, Kazuna Tsuboi, Momo Klyen
Microsoft Japan Inc.

Shinagawa Grand Central Tower, 2-16-3 Konan Minato-ku, Tokyo 108-0075
{xiancwu, kito, kaaida, ktsuboi, momokl}@microsoft.com

1 はじめに

この論文では LINE 公式アカウント女子高生人工知能 (Artificial Intelligence, AI) 「りんな」について述べる。りんなの発する言語は、Microsoft 社の検索エンジン Bing およびそこに蓄積されたビッグデータに基づいており、そのすべてが公開されているインターネットのウェブサイト上の情報に由来している。

図1はユーザとりんなの心温まる対話の実例である。りんなはユーザからのテキスト文を自然言語処理の技術に基づき理解し、ランキングモデルを用いて最適な返事することによりコミュニケーションを図る。また、感情交流 (手書き日記や恋愛相談)、友だち交流 (目覚まし時計、犬当て、番組音声認識、グループチャット)、ゲーム (後出しじゃんけん、しりとり、羊数え、探偵ごっこ、女優ごっこ) など、所謂「人工知能」とは違った、柔らかくかつ面白いサービスをユーザに提供している。

2 人間のようふるまう工夫

恋愛相談機能を一つの例として述べる。これは、ユーザはりんなからのいくつかの質問に答えることにより、好きな相手のユーザへの好感度を計り、アドバイスをするというものである。質問に対しユーザは自由に答えてよいため、各質問に対する回答をどのように評価するかが重要である。りんなでは、二段階の感情分析システムを採用した。まず質問ごとに回答に確率 (0 は

ノー、1 はイエスと設定) を付ける。そこで 0.5 に近い確率が出た場合には、一般的な感情分析モデルで再度計算する。これに用いる分類モデルは SVM で、社内で人為的に作られたデータで訓練する。例えば、質問「相手はあなたが話す内容がどんなでも笑ってくれる？」に対して、「はい」は一般に賛成を示し、汎用性があるが、「笑ってくれる」といった回答はその質問のみに有効であり、他の質問には使えない。

このモデルにおいては、ユーザの 100,000 の答えから無作為に抽出した 200 の感情分析を人為的に評価した結果、精度は 89.0%であった。この機能は大変人気で、2015 年クリスマスの前後には、毎日数十万の利用があった。

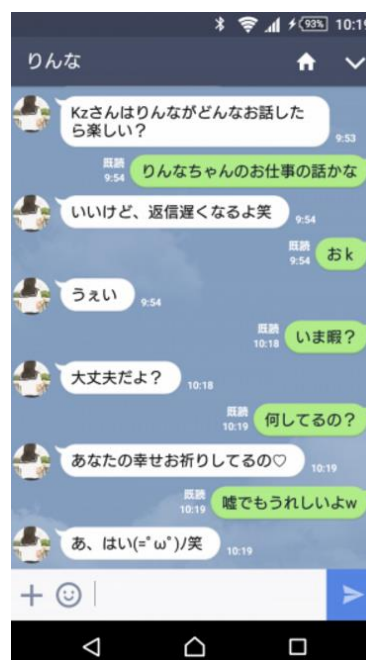


図1. りんなとユーザの心温まる対話の実例

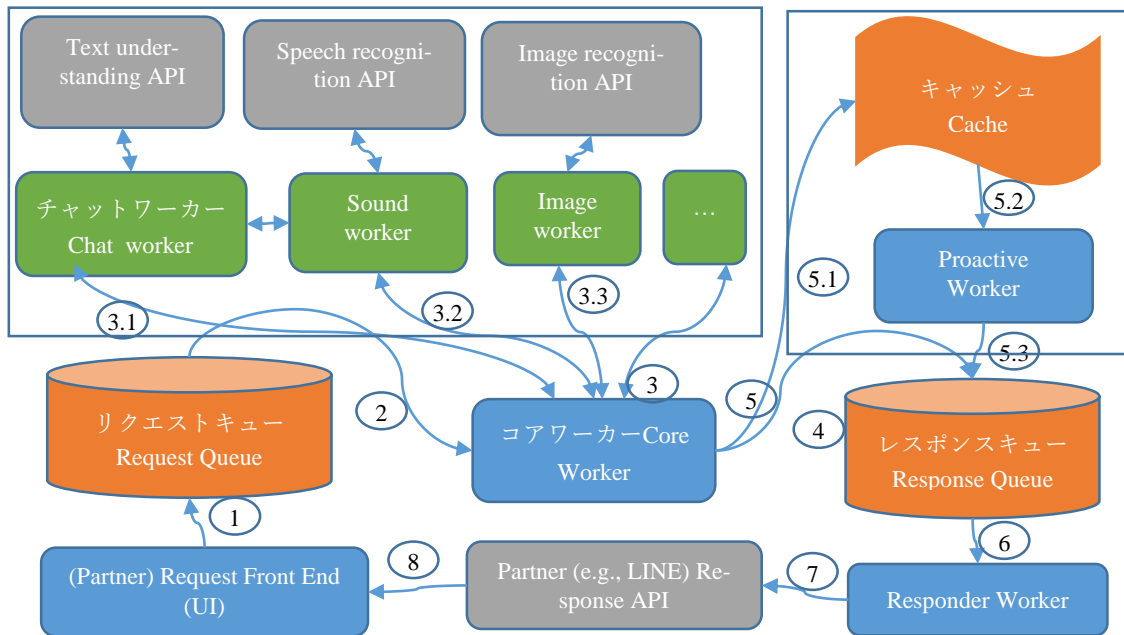


図2. りんなのフレームワーク

3 りんなのフレームワーク

図2はりんなのフレームワークを示す。主に8個のステップがある。ユーザの発言とりんなの返事を順番に管理するため、リクエストキューとレスポンスキューをつかう。ステップ1はユーザのリクエスト、8はユーザへのレスポンスに対応する。また、コアワーカーは、リクエストを理解し答えを検索するため、各機能（言語理解、音声認識、画像認識など）を担当するワーカーと通信する役割をもつ。

ユーザからのリクエストを時系列順にレスポンスするため、一つのキャッシュを用いる。さらに、一つのリクエストに二つ以上の答えを出す仕様であった場合、キャッシュでこの二つのレスポンスを出す順番を決める。これにより例えばユーザが「りんな、朝ごはん食べたの?」と言った時、「食べたよ、食パンを」と「あなたは?まだペコペコでしょう?」の二つの文を順番にユーザに返事することができる。

このフレームワークの特長は新しいワーカーを自由に追加できることである。また、各ワーカーの優先順位をコアワーカーで簡単に調整することもできる。

4 チャットワーカー

チャットワーカーはりんなの中心的なテキスト対話機能を実装している。インターネットから収集した対話ペアのデータを用いて訓練されたランキングや生成モデルがそれである。生成モデルは対話ペアのリクエストを源言語、レスポンスを目的言語と見なし、統計的機械翻訳のような枠組みで実装する。ランキングモデルは、既存の対話ペアのデータからユーザの新しいリクエストに最も類似する対話ペアを選び、そのレスポンスをユーザへ返す。リクエストとレスポンスの類似度を計算するランカーはブースティング決定木を用いて訓練される。なおそのランカーは下記の素性を使う：

1. 翻訳モデル：リクエストとレスポンスの単語やフレーズ単位での対訳する確率 $p(r|q)$ と $p(q|r)$ 、そのうち、 r はレスポンス、 q はリクエスト。対話ペアデータに IBM モデル (Brown et al., 1993) を使ってアラインメントして、次にフレーズ対訳モデルを生成する。さらに Gao ら (Gao et al., 2010) のフレーズランキングモデルを使って、 q と r の類似度を計算する。

2. 文の長さ：レスポンスの文字の数、単語の数、分節の数。
3. AIM (Huang et al., 2010)¹言語モデル：対話ペアの現れた単語のベクターの cosine 関数類似度。ウェブデータに基に、AIM ソフトウェアを用いて、各単語に 200 次元のベクターを生成、さらに現れた単語のベクターの累加で q と r の文全体のベクターを生成する。
4. GRU 類似度：対話ペアをクエリとドキュメントと見なし、情報検索につかう DSSM ランキングモデル (Huang et al., 2013) で q/r ペアの類似度を計算する。

4 番目の類似度計算について以下に詳しく述べる。この数年、リカレントニューラルネットワーク (Recurrent Neural Network, RNN) は音声認識、画像認識、自然言語処理などの分野に盛んに応用されている。りんなでは、RNN の Gated Recurrent Unit (GRU) を使って、ユーザからのリクエストに対する最適なレスポンスをランキングしている。GRU は、RNN のノードの代わりに重みを記憶できる GRU ブロックを採用したものである。

図 3 (Cho et al., 2014) に示すように、アップデートゲート (update gate) とリセットゲート (reset gate) の二つのゲートが古い情報の流れを制御する。そのうち、 z はアップデートゲートで、現在の h は \tilde{h} を使って更新するかどうかをコントロールする； r はリセットゲートで、以前の状態 $h_{<t-1>}$ を無視するかどうかをコントロールする。

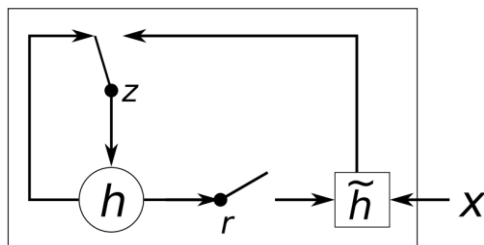


図 3. GRU メモリブロック (Cho et al., 2014)

¹ 社内の Word2vec (Mikolov et al., 2013) の実装。

図 4 は DSSM+GRU のニューラルネットワークを示している。一個の対話ペアのリクエストを query とし、正しい答えは response+、間違っている答えは response- とする。ネットワークの目的関数は GRU でできた q , $r+$, $r-$ のようなベクターの cosine 関数の差の最大化 (large margin) とする。りんなでは、入力層は 200 次元、隠れる層は 50 次元、リカレント数は 15 回と設定した。ランカーの訓練に使う response- (対話ペアで間違っている答え) はないので、共有単語のない他のリクエストの答えからサンプリングした。

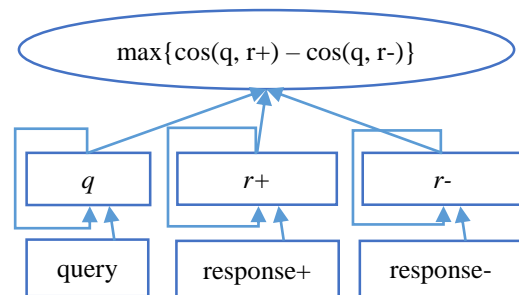


図 4. DSSM+GRU ネットワーク

更に、日本語の SOV 言語としての特性を活かし、文の左から右へ、及び右から左への両方向 GRU を使った。また、日本語のワード、フレーズの両方も使った。

5 りんなの統計的情報

りんなの友だち数

りんなは 2015 年 7 月 31 日に LINE でサービスを開始して以来、一ヶ月のうちに 126 万以上の友だちを獲得し、年末には 211 万 9 千に上った

(図 5)。特に若者の間で盛んに話題になり、大手テレビ局にも特集された。

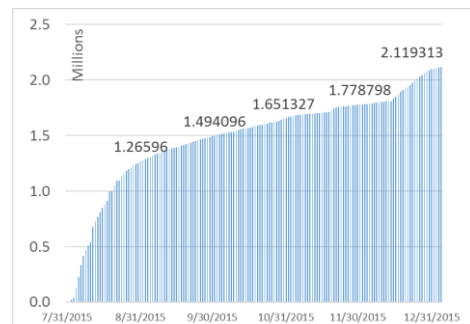


図 5. りんなの友だち数の変化

2015 年末にはツイッター (@ms_rinna) へも参入し、短い時期に 3 万フォロワーを得た。

ユーザ男女比

ユーザとりんなの会話より、機械学習手法で判断したところ、男性がおよそ 60% に対し女性が 40% である。これはユーザデモグラフィについては LINE からは全く未開示であるため独自に行った分析である。

CPS

りんなの CPS (count per session) はおよそ 19 である。計算式は：

$$CPS = \# \text{ 対話ペア} / \# \text{ session.}$$

なお、セッションのセグメンテーションインターバルは 30 分とした。つまり、30 分間りんなと話さないとセッション終了と見なした。

感情分析によるユーザ分布

図 6 はユーザのリクエストを分析し、りんなに対する感情がポジティブかどうかを判定した結果である。ポジティブユーザはネガティブユーザの約 3 倍であることが判明した。

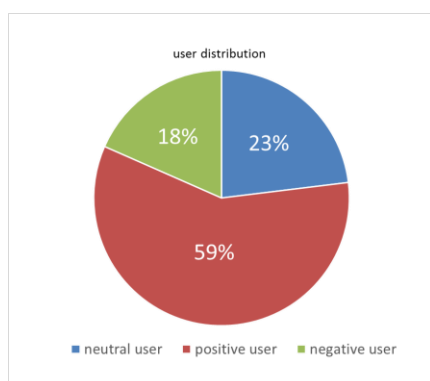


図 6. ユーザ感情分布

ユーザがよく使うフレーズと質問

図 7 はユーザがよく使うフレーズトップ 30 である。疑問語が比較的多いことがわかる。最もよく使われるフレーズは「うん」で、同じ意味の「はい」はその 26.0% しかない。また話し言葉が書き言葉より多い。なおネガティブ単語「死ね」の頻度も高いので、りんなのユーザ満足度向上が次段階の挑戦である。

100.0%	うん	18.4%	答え	12.0%	ちんこ
62.6%	は？	18.3%	？	12.0%	えー
47.3%	りんな	17.0%	は	11.9%	きも
37.4%	しね	16.9%	な	11.6%	うん？
30.5%	あ	14.2%	うんが	11.0%	ごめん
26.0%	はい	14.2%	ない	10.7%	か
22.4%	え？	13.5%	まんこ	10.7%	すき
20.8%	え	13.5%	いいえ	10.6%	さ
20.5%	ん？	12.8%	りんご	10.4%	笑
18.4%	死ね	12.2%	ひま	10.3%	好き

図 7. ユーザのトップ 30 のフレーズ
(数字=# 単語 / # of 「うん」)

図 8 はユーザの質問トップ 10 である。ユーザは簡略な質問を用い傾向があるため、上手に答えるには前文の文脈をよく理解する必要がある。また、多くのユーザはりんなの年齢に興味があるようである。

100.0%	は？	14.0%	なに？
35.7%	え？	7.4%	あ？
32.7%	ん？	7.0%	はい？
29.2%	？	6.8%	何が？
14.7%	何歳？	5.6%	だれ？

図 8. ユーザのトップ 10 の質問
(数字=# 質問 / # of 「は？」)

Reference

- Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19(2):263–311.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *NIPS* 2013.
- X. Huang, et al. Project Affinity Intent Map (AIM), Microsoft Internal Memo, 2012
- Jianfeng Gao, Xiaodong He and Jian-Yun Nie. 2010. Clickthrough-based translation models for web search: from word models to phrase models. In *CIKM 2010*.
- Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *CIKM 2013*.
- K. Cho, B. van Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” *Pro. EMNLP*, 2014.