

LINE、音声処理における世界最大規模の国際学会「INTERSPEECH 2021」にて国内トップクラスの6本の論文採択

2021.08.30 技術情報

LINEとNAVERとで14本の論文採択

非自己回帰型音声認識の新方式および音声合成における句境界予測の研究での評価

LINE株式会社（本社：東京都新宿区、代表取締役社長：出澤 剛）は、音声処理における世界最大規模の国際学会「INTERSPEECH 2021」にて6本の論文が採択されましたので、お知らせいたします。

「INTERSPEECH」は、International Speech Communication Association（ISCA）が主催する音声処理における世界最大規模の国際会議です。2021年は22回目の開催となり、約2000件の投稿から約1000件の論文が採択されました。採択された論文は本日から9月3日にかけてオンラインで開催される「INTERSPEECH 2021」にて発表を行います。

■非自己回帰型音声認識の新しい方式、音声合成における句境界予測に関する研究などが評価され採択

今回の「INTERSPEECH 2021」では、まず音声認識の研究としては、高速な音声認識を実現する手法として注目を集めている非自己回帰型音声認識^{*1}の性能向上に関する論文 [1] が採択されました。本論文では複数のTransformer^{*2}層とCTCを^{*3}を併用した既存の手法の課題である、各トークン（文字や単語の単位）を考慮できない点の解決に取り組みました。通常の音声認識ではTransformerの最終層の出力を用いて認識を行います。本論文ではそれに加えて中間層から得られる特徴を用いて仮の認識を行い、その認識結果を後続のTransformer層にフィードバックする方式を提案しました。中間層で一度認識を行うことで、後段のTransformer層にトークン領域の情報を伝えることができ、トークン間の関係性を考慮した特徴を出力することができます。この方式は既存手法の簡単な拡張でありながらも、他の最新手法を上回る精度を示しました。

また、音声合成の研究としては、音声の適切な位置に無音区間、ポーズを挿入することで合成音声の品質を向上させる句境界予測の研究 [3] が採択されました。提案法では、近年言語処理分野の多くのタスクで活用されるBERT（Bidirectional Encoder Representations from Transformers）の表現力に着目し、BERTと従来手法であるLSTM（Long Short Term Memory）を組み合わせることで句境界予測の精度が大きく改善し、かつ合成音声の品質も向上しました。またNAVERと共同で進めていた、「ICASSP 2020」で発表した Parallel WaveGAN (PWG) ^{*4}をより高品質にするための取り組みである、Multi-band harmonic-plus-noise PWG の研究 [4] が採択されました。古典的な信号処理手法に着想を得て、PWGに周波数帯域ごとの音声の周期・非周期成分を重み付けて混合させる機構を導入することで、より高い品質の音声合成を可能にしました。

^{*1} 非自己回帰型音声認識：過去に生成したテキストに依存せずに、各時点の音声を認識する手法

^{*2} Transformer：ニューラルネットワークのモデルのひとつであり、過去の出力に次の出力が依存する再帰型ニューラルネットワーク(RNN)を用いることなく、Attention機構（各トークンごとの関係性を学習する機構）を用いて時系列のデータを並列的に扱うことが可能なモデル

^{*3} CTC：音声データとテキストのように、ニューラルネットワークの入力の系列長が出力の系列長よりも大きい場合に用いられる学習手法

^{*4} Parallel WaveGAN (PWG)：機械学習の生成モデルのひとつであり2つのニューラルネットワークを用いて学習を行って入力されたデータから新しい擬似データを生成する「敵対的生成ネットワーク（Generative Adversarial Network / GAN）」を用いた非自己回帰型音声生成モデル

採択された論文

1. Jumon Nozaki and Tatsuya Komatsu,"Relaxing the Conditional Independence Assumption of CTC-based ASR by Conditioning on Intermediate Predictions".
2. Tatsuya. Komatsu, Shinji. Watanabe, Koichi. Miyazaki, and Tomoki. Hayashi, "Acoustic Event Detection with Classifier Chains".
3. Kosuke Futamata, Byeongseon Park, Ryuichi Yamamoto and Kentaro Tachibana, "Phrase break prediction with bidirectional encoder representations in Japanese text-to-speech synthesis".
4. Min-Jae Hwang, Ryuichi Yamamoto, Eunwoo Song and Jae-Min Kim, "High-fidelity Parallel WaveGAN with Multi-band Harmonic-plus-Noise Model".
5. Yu Nakagome, Masahito Togami, Tetsuji Ogawa and Tetsunori Kobayashi, "Efficient and Stable Adversarial Learning Using Unpaired Data for Unsupervised Multichannel Speech Separation".
6. Masahito Togami and Robin Scheibler, "Sound Source Localization with Majorization Minimization".

【LINEが注力する基礎研究について】

LINEでは、AI事業を戦略事業の一つとして位置付け、AI技術の研究・開発およびAI技術を活用した事業の発展を加速させることを目的に、NAVERとの連携も行いながら、新たなAI関連サービス・新機能の創出を進めるとともに、それらを支える技術の基礎研究に注力しています。そして、「データ基盤開発」「データ分析」「機械学習」「AI技術開発」「基礎研究」を担う各チームが事業や担当領域を超えて連携することで、「研究＞開発＞事業化」のサイクルをスピードアップすることを目指しています。

基礎研究においては、機械学習を軸に、音声処理、言語処理、画像処理などに注力しています。2021年2月には、音声・音響信号処理分野で世界最大の国際学会「ICASSP」にて国内トップクラスの7本の論文^{*5}、2021年7月には、コンピュータービジョン分野における世界最高峰の国際会議のひとつ「ICCV 2021」にて2本の論文^{*6}が採択されるなど、着実な成果を収めております。

^{*5} 2021年2月26日発表プレスリリース：<https://linecorp.com/ja/pr/news/ja/2021/3639>

^{*6} 2021年7月28日発表プレスリリース：<https://linecorp.com/ja/pr/news/ja/2021/3843>

【LINEが提供するAIテクノロジーブランド「LINE CLOVA」】

LINEが提供するAIテクノロジーブランド「LINE CLOVA」は、さまざまなAI技術やサービスを通して、生活やビジネスに潜む煩わしさを解消すること、社会機能や生活の質を向上させることで、より便利で豊かな世界をもたらすことを目指しています。現在、音声に関連する技術として「CLOVA Speech（音声認識）」「CLOVA Voice（音声合成）」、そしてそれらの技術を組み合わせたソリューションの提供も行っています。

たとえば「LINE AiCall」は、「CLOVA Speech」と「CLOVA Voice」、および会話制御の仕組みを組み合わせることで、ユーザーの要望に対してAIによる自然な対話応答を実現し、目的を達成するソリューションで、現在行政や飲食店舗、各種コールセンターなどでの活用が進んでいます。また、昨年発表した「CLOVA Note」は、会議における会話を

高い精度で認識し、議事録として記録・管理するためのアプリケーションです。長時間の録音データ解析に特化した音声認識モデルを用いることで、高い認識精度を実現します。

今後も、AI技術に関連する基礎研究を積極的に推進することで、既存サービスの品質向上や、新たな機能・サービスの創出に努めてまいります。

LINEでは、今後も事業発展やサービスの価値向上のための取り組みを積極的に行い、コミュニケーションインフラとして更なる成長・拡大の可能性を広げてまいります。
