経営情報学科

キーワード：自然言語処理、語彙意味論、言語モデル、文生成

教授 / 博士（国際情報通信学）
谷垣　宏一
Koichi Tanigaki

学歴

東北大学 工学部 情報工学科、東北大学 大学院 情報科学研究科 情報基礎科学専攻 博士前期課程、早稲田大学 大学院 国際情報通信研究科 国際情報通信学専攻 博士後期課程

経歴

三菱電機株式会社 情報技術総合研究所 首席研究員、国際電気通信基礎技術研究所（ATR）研究員

相談・講演・共同研究に応じられるテーマ

蓄積されたテキストデータ・文書アーカイブや、情報システムログなど、数量化・集計が自明でないビッグデータの活用に関する共同研究、技術相談

メールアドレス

tanigaki@fukui-ut.ac.jp

主な研究と特徴

「多重文脈における文章生成の深層学習モデル」

医師や技術者などの専門家が診断や処置の内容を記載する報告書は、業務上作成が義務付けられる重要な文書であるが、その作成は、繁忙な専門家にとって負荷の大きなサブタスクとなっており、自動化の実現が期待される。そのような専門家が作成する報告書の一例として、ソフトウェア技術者が作成するソースコード修正説明文（コミットメッセージ）を対象に、説明文の自動生成方式を開発し、実証実験を行った。

有益とされる説明文を構成するには、直接的な説明対象であるコード修正部分（＝第1の文脈）のサマリに加えて、修正に至った理由（＝第2の文脈：バグ報告や機能改善要望などとの関連）についても簡潔な説明を加える必要がある。これまでの先行研究では、機械翻訳や文書要約で用いられる系列変換モデルを適用してコード修正部分から説明文を生成（変換）しており、このため通常、コードからは読み取り困難な修正理由については、十分な説明を生成できないという課題があった。そこで本研究では、マルチエンコーダ型のニューラルネットワークを構成して、性質の異なる複数の文脈を同時入力し、注意機構を用いて参照する文脈を動的に切り替えながら説明文を生成する深層学習モデル（図1）を開発した。インターネット上のリポジトリサービスから18万件のデータを入手して実施した学習実験の結果、先行研究を上回る精度で文章が生成され、本方式の有効性が実証された。

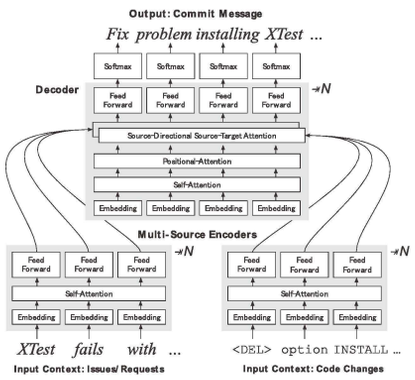

図1. 多重文脈における文章生成の深層学習モデル

「階層ベイズモデルによる語義とメタファーの認識」

人間の豊かで円滑な言語コミュニケーションは、しばしばメタファー（比喩、言葉のあや）によって彩られており、字義通りの意味からは飛躍した意味での解釈が期待される。メタファーを正確に認識する技術は、今後、対話ロボットなどの応用AIが、我々の生活やコミュニケーションに介在して円滑に機能する上で重要な要素技術である。メタファーの中でも慣用的に用いられるメタファーの認識は、語義曖昧性解消問題の一部を構成しており、自然言語処理において歴史の長いタスクとして取り組まれている。

本研究では、対象語を限定しない語義曖昧性解消（all-words WSD）のための新しい教師なし学習モデルを提案した。all-words WSDは、辞書知識を言語処理に活用する基礎技術として実用化が期待されるが、扱う語義の種類が膨大で、かつ分布がドメインに強く依存する性質があるため、ラベル付きコーパスの構築を前提とする教師あり学習では実用化を見込むことが難しい。そこで本研究では、ラベルなしコーパスに出現する種々の語と膨大な語義の間に自然な対応を推定するため、2つの制約をモデル化した。1)同じ語の各出現における語義は、単語タイプ毎の事前分布に従う。2)類似した文脈に出現する種々の語の語義は、各語の語義割り当てを平滑化して得られる分布に従う。これら2つの制約を階層ベイズモデル（図2）によって同時適用することで、教師なしall-words WSDを実現する。ベンチマーク・データセットを用いた実験結果より本手法の有効性が示された。

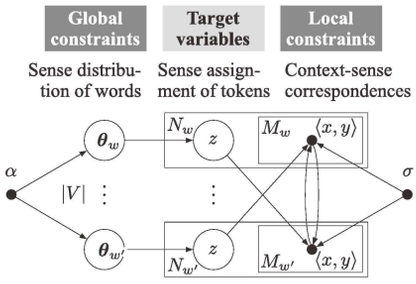

図2. 語の出現と語義の対応の確率的生成モデル

今後の展望

上述のように説明文生成においては、マルチエンコーダ型文章生成モデルを提案した。本モデルは複数の文脈から必要な情報を抽出して文章を構成しようとするモデルであり、基本的な考え方は多くの説明文作成タスクに共通と期待される。今後、種々のタスクへの適用と効果検証が待たれる。一方、ニューラル文生成の一般的な特徴として、流暢性は高いが妥当性が低い文が生成されることが知られており、例えば機械翻訳においては、原文の一部が訳文に表出されない、原文にはない内容が訳文に挿入される、といった問題が発生する。こうした問題は特に、正確性が強く求められる医療などの分野向けには深刻な課題であり、エラーを抑制する機構（例えば敵対的生成ネットワークの枠組みを利用するなど）が必要である。メタファー認識においては、GlobalとLocal, 2つの確率的制約により階層ベイズモデルを構成し、語義識別タスクにおいて有効性を実証した。しかし、辞書語義の範疇を超えて創造的なメタファーも解釈するには、より広範かつ多様な知識に基づく推論が必要であり、言語学的・認知心理学的な知見を反映した数理モデルの確立に取り組んでいる。日本語のメタファー研究においては、大規模なラベル付きコーパスが存在しないことが課題であり、日本語コーパスの早急な整備拡充が待たれる。

所属学会

一般社団法人 情報処理学会会員　（平成9年～現在まで）
一般社団法人 言語処理学会会員　（平成23年～現在まで）
一般社団法人 人工知能学会会員　（平成29年～現在まで）

主要論文・著書

谷垣宏一、撫中達司、匂坂芳典、語の出現と意味の対応の階層ベイズモデルによる教師なし語義曖昧性解消、情報処理学会論文誌, Vol. 57, No. 8, pp. 1–11（2016）.
谷垣宏一、撫中達司、匂坂芳典、文脈と意味の対応密度最大化による教師なし語義曖昧性解消、情報処理学会論文誌, Vol. 57, No. 3, pp. 1-11（2016）.
Tanigaki, K., Shiba, M., Munaka, T. and Sagisaka, Y.: Density Maximi-zation in Context-Sense Metric Space for All-words WSD, Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics（ACL2013）, Vol. 1, pp. 884–893（2013）.