

Title	文書ビッグデータから含意を認識する手法の開発
Sub Title	A method of textual entailment from big data
Author	斎藤, 博昭(Saito, Hiroaki)
Publisher	慶應義塾大学
Publication year	2019
Jtitle	学事振興資金研究成果実績報告書 (2018.)
JaLC DOI	
Abstract	本研究における含意とは、ある文（文章）Tと仮説Hという二つの文において、「Tが正しい時にHも正しいと推定できる関係」である。含意関係認識をコンピュータに行わせるタスクは十年以上前から行われてきたが、昨今の人工知能研究の発達に伴い、研究手法の新たな開発が活発になっている。今回は、日本語を対象として深層学習を用いた。深層学習の全体の枠組みは踏襲したが、その構成部品として従来行われてきた単純なRNNやLSTMを使うのではなく、2文間の文節対応を考慮し、かつ外部メモリを取り入れることとした。似た手法は英語の含意関係認識では試みられているが、日本語では初である。具体的には、TとHをまず分散表現（ベクトル）に変換し、それぞれについて外部メモリ付LSTMでノードの値を求めた後、アラインメントを施し、含意度を計算するというものである。文節対応においては一般の2分木モデルとともに日本語における自由な語順や係り受けの不安定さに対応できる多子木モデルを試した。さらに、さまざまなハイパーパラメータを探索した。 結果として、単純な深層学習モデルの性能を上回り、文節関係や外部メモリの効果、および多子木モデルによる精度向上は確認できた。しかしながら、人手で素性を作りベクトル化する統計ベースの手法には及ばなかった。深層学習では多くのデータが必要だが、今回は1000に満たないデータしか用意できず、結果としてテストデータに含まれる多くの未知語への対応が不十分となった。データ量が桁違いに多い英語に関してデータ量と精度との関係を調べたところ、やはりデータ量に比例して精度が上がるのが確かめられたので、含意関係認識という複雑なタスクを深層学習で攻めるには、データ量が少なかったというのが反省点である。 Recognition of textual entailment (RTE) is a research field in natural language processing, where a relation between two sentences is detected; a text T entails a hypothesis H when it is possible to infer "H is true" from "T is true." Various methods have been developed to this task and deep learning is actively pursued recently. In this research tree-LSTM with external memory is adopted. The proposed method surpassed the baseline, but did not reach the performance of conventional support vector machine approaches. The reason was probably the shortage of Japanese RTE data. This result implies that deep learning for complicated tasks needs huge good data.
Notes	
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=2018000005-20180034

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

研究代表者	所属	理工学部	職名	准教授	補助額	200（B）千円
	氏名	斎藤 博昭	氏名（英語）	Hiroaki SAITO		
研究課題（日本語）						
文書ビッグデータから含意を認識する手法の開発						
研究課題（英訳）						
A Method of Textual Entailment from Big Data						
1. 研究成果実績の概要						
本研究における含意とは、ある文（文章）Tと仮説Hという二つの文において、「Tが正しい時にHも正しいと推定できる関係」である。含意関係認識をコンピュータに行わせるタスクは十年以上前から行われてきたが、昨今の人工知能研究の発達に伴い、研究手法の新たな開発が活発になっている。今回は、日本語を対象として深層学習を用いた。深層学習の全体の枠組みは踏襲したが、その構成部品として従来行われてきた単純なRNNやLSTMを使うのではなく、2文間の文節対応を考慮し、かつ外部メモリを取り入れることとした。似た手法は英語の含意関係認識では試みられているが、日本語では初である。具体的には、TとHをまず分散表現（ベクトル）に変換し、それぞれについて外部メモリ付LSTMでノードの値を求めた後、アラインメントを施し、含意度を計算するというものである。文節対応においては一般の2分木モデルとともに日本語における自由な語順や係り受けの不安定さに対応できる多子木モデルを試した。さらに、さまざまなハイパーパラメータを探索した。 結果として、単純な深層学習モデルの性能を上回り、文節関係や外部メモリの効果、および多子木モデルによる精度向上は確認できた。しかしながら、人手で素性を作りベクトル化する統計ベースの手法には及ばなかった。深層学習では多くのデータが必要だが、今回は1000に満たないデータしか用意できず、結果としてテストデータに含まれる多くの未知語への対応が不十分となった。データ量が桁違いに多い英語に関してデータ量と精度との関係を調べたところ、やはりデータ量に比例して精度が上がる事が確かめられたので、含意関係認識という複雑なタスクを深層学習で攻めるには、データ量が少なかったというのが反省点である。						
2. 研究成果実績の概要（英訳）						
Recognition of textual entailment (RTE) is a research field in natural language processing, where a relation between two sentences is detected; a text T entails a hypothesis H when it is possible to infer "H is true" from "T is true." Various methods have been developed to this task and deep learning is actively pursued recently. In this research tree-LSTM with external memory is adopted. The proposed method surpassed the baseline, but did not reach the performance of conventional support vector machine approaches. The reason was probably the shortage of Japanese RTE data. This result implies that deep learning for complicated tasks needs huge good data.						
3. 本研究課題に関する発表						
発表者氏名 (著者・講演者)	発表課題名 (著書名・演題)	発表学術誌名 (著書発行所・講演学会)	学術誌発行年月 (著書発行年月・講演年月)			