

0cm

要素技術
2cm

本

章では,本研究に関連する要素技術について説明する.

0cm

自然言語処理に関する要素技術
2cm

自

然言語の単語や文を計算機上で表現するための分散表現獲得手法に

どの自然言語において意味を持つ最小の単位のことであり、文を形態素に分割し、各形態素の品詞などを判定する技術を形態素解析という。英語の文では、単語と単語の区切りがほ

得する。前者の手法を Skip-gram

といい、後者の手法を Continuous Bag-of-Words (CBOW) という。

Doc2Vec

[?]は Word2Vec をベ-

ス

とし

た、

文書

をベ

クト

ル空

間上

に写

像し

て分

散表

現を

得る

自然

言語

処理

の手

法で

ある。

Para-

graph

ID

0cm

At-
ten-
tion
2cm

機

機械翻訳のタスクに対して考案された,
LSTM

を用いる
Encoder-
Decoder

モデルは可変長の文を固定長のベクトルにエンコードするため、長い入力文になるほど隠れ層のノード数が不足し、

0cm
Trans-
form-
ers
2cm
Transformer

[?]は他モデルで頻繁に用いられてきたRNNを用いずにAttention機構のみを基本構造とするEncoder-Decoderモデルである。図??にその概略図を示す。Transformerのエンコーダおよびデコーダは

0cm
BERT
2cm
Bidi-
rec-
tional
En-
coder
Rep-
re-
sen-
ta-
tions
from
Trans-
form-
ers
(BERT)
[?]
は、
2018
年
に
Google
が
発表
した
言語
モデ
ルで
あり、
複数
の双
方向
Trans-
former
に基
づく
汎用
言語
モデ
ルで
ある。
これ
まで
の言
語モ
デル
は特
定の
学習
タス

0cm
画像処理に関する要素技術
2cm
画
像処理に関する要素技術について説明する.
0cm
VGG
2cm
VGG16
[?]とは、ImageNet
と呼ばれる大規模画像データセットで学習済みの畳み込みニューラルネットワーク