

ネット出品商品説明 生成支援API

魏 琪¹・趙 一平²

¹ HQ Lab (〒179-0074 東京都練馬区春日町4-25-18)

E-mail:qi.wei@hqlabo.com

² 株式会社iWing (〒192-0046 東京都八王子市明神町2-26-9 MZビル301C)

E-mail:yiping.zhao@iwingcorp.com

近年、メルカリなど一般消費者も利用できる出品アプリの応用に伴いて、一般消費者も簡単に商品売れるようになる。出品初心者もしくは大量出品者は商品説明の作成は苦勞にする。私たちは商品説明自動生成APIを開発し、出品者の商品説明生成を支援する。

1. はじめに

近年、ブロードバンドネットワークの普及により、一般消費者もインターネットを用いたネットワーク型の取引を容易に利用できる。例えば、今主婦たち大人気のメルカリでは、一日50万点以上のアイテムが出品されていることを報告された。しかし、出品のアイテム、特に人気カテゴリ中には、商品の説明は人気度を大きく左右される。

すぐに売れる商品説明の書き方ガイドがあるのですが、出品の初心者は商品の説明は苦勞になる。または、大量出品者は大量の商品説明の書くも面倒な作業になる。

近年、野球新聞自動生成[1]、株市場報告自動生成[2]などテキスト自動生成技術を成功に応用する。しかし、商品説明生成について、現在テンプレートに穴埋めを行う手法を用いた生成モデルが多い。本研究は、seq2seqモデルを商品テキスト自動生成を成功する。出品システムの利用者に支援するの目的である。

2. テキスト生成API

本研究のシステム構造は図1で示す。システム利用者は写真またはキーワードを入力して、

商品説明生成支援システムにパスする。そして、本システムはまず、キーワード生成インターフェイスを経由、写真中の商品情報抽出して、または入力したキーワードを整形して、テキスト生成システムのインプットとして使う。テキスト生成システムから処理して、商品の説明をアウトプットして、利用者に見せる。

テキスト生成システムは入力前処理ユニット、テキスト生成ユニットとテキスト後処理ユニットと三つユニットを含める。それぞれユニットを経由し、最後に商品説明を作成する。

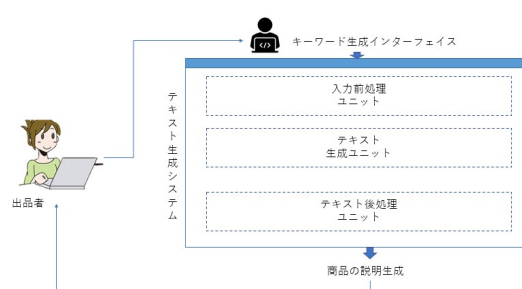


図1 商品説明生成支援システム

図2は商品説明APIのサンプルになる。入力をJSONに変換し、商品説明APIで計算して、テキストを生成する。図3はAPI経由、3回生成の結果である。

2.1 データ収集

本研究は、インターネットショッピングウェブサイトから、男性、女性とキッズ服の写真と商品説明20万件ぐらい収集した。収集したデータは大、中、小分類で分ける。大、中分類は表1と表2で示す。収集したデータはクリーニングを行う。商品説明無関係な情報、例えば、店の紹介や、産地などを除いている。続いて、ルールベースで、それぞれの商品特徴と商品説明文のペアを抽出する。この抽出データはテキスト生成モデルのトレーニングとして使う。

No.	大分類
1	レディース
2	メンズ
3	ガールズ
4	ボーイズ

表1 大分類リスト

No.	中分類
1	トップス
2	ワンピース
3	コート
4	パンツ
5	水着
6	アクティブウェア

表2 中分類リスト

2.2 テキスト前処理

本研究はキーワード生成インターフェイスからもらったキーワードを事前準備したパターンのテキストにランダムに埋め込んで、生成モデルの入力として使う。

2.3 テキスト生成モデル

本研究はNMT (Neural Machine Translation) [3] モデルを用いて、テキスト生成を行う。NMTは深層学習技術を機械翻訳に適用したニューラル機

械翻訳モデルです。本研究は単一言語翻訳と考えて、NMTを利用した。本研究のモデルは図3を示す通り、注意機構を利用したseq-to-seqモデルになる。入力単語を埋め込み表現に変更し、隠れ層に経由、そして、注意機構をかけて、投射層に入れ、最後はSoftmax関数をかけて、出力する。

本研究はWord2Vec[4]を用いて、単語の埋め込み表現に変換する。Word2VecはTomas Mikolovらが考えてニューラルネットワークを用いて、skip-gramモデルと呼ばれた言語モデル。単語の分散表現の計算手法である。入力単語の前後単語を予測するようにニューラルネットワークを学習される。学習後、文章中の単語を任意次元のベクトルに変換し、意味的に似ている単語は空間上の近い位置に配置され、単語同士の演算や単語の類似度の導入を可能である。本研究は単語の埋め込み表現に変換として行う。

本研究は2.1から収集のデータセットをMcCab[5]を使って、単語を分ける。Word2Vec訓練データセットを作成する。作成したデータセットWord2Vecモデルに学習する。

2.4 テキスト後処理

入力テキストはテキスト生成モデルに入れて、5本のテキストを作成する。5本のテキスト中、ルールベースで、関連性低いテキストを除外する。そして、全部のテキストの類似度を計算し、重複テキストを削除する。最後、テキストの可読性を向上させるのため、フィルターを掛ける。

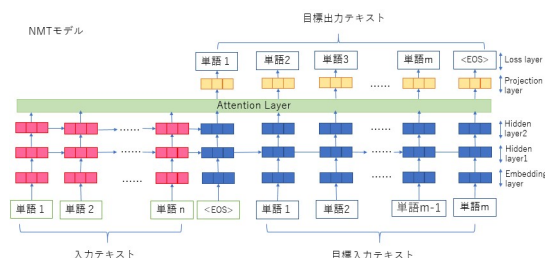


図3 seq2seqモデル

3. システムの評価

商品説明のテキスト生成の評価について、自動評価が難しいので、今回人工評価になった。本研究は商品の特徴（素材、利用季節、袖などの特徴、色、サイズ）合う度と言葉通じ度の二方面から評価する。各項目の値は0~1になる、値高いほどがいい。評価はレディースブラウザ、レディースTシャツとメンズトレーナー三種類の商品カテゴリから、商品10件ずつランダムに抽出、一件の商品10回ずつテキストを生成して、生成の結果を人工的に評価する。

商品の説明自動生成テキストの評価は表3で示す。商品の特徴の説明は全項目0.75以上の点数取った。特に、色とサイズの正しさが高い。素材について、複数の素材と比率入力した場合は、生成したテキストが素材の比率が違う、素材が足りないのケースが多い傾向が見える。

言葉通じ度について、全体の点数は0.47になった。テキストの前後組み合わせや複数似ている意味説明の混在や、言葉自身おかしいのことが大きな原意と見られる。その中、言葉自身おかしいの一番原因はトレーニングデータセットはダミーデータが含まれる。インターネットから収集したデータには、日本語母語話者だけのデータを限らないのため、データのクリーニングが必要と考える。

4. まとめと今後の展望

本研究は商品説明生成支援の目的で、テキスト自動生成技術を用いて、テキスト生成APIを開発した。人工評価の結果、商品の特徴関連の内容生成はいい評価が貰った。商品特徴以外の説明も含めるので、評価者にいい印象を与えた。

一方、言葉通じはまだ課題が残っている。今後、商品説明トレーニングデータのクリーニングがもっと力が入ると考える。テキストの順番と繋がる言葉機能も含めると予定する。

参考文献

- 1) 工藤 健太郎,大川 恭平,金澤 慧,村井 源 WPAを用いた野球の戦評の自動生成情報知識学 0917-1436 情報知識学 2019-29-2,181-186
- 2) 酒井浩之,坂地泰紀,和泉潔,松井藤五郎,入江圭太郎 経済テキストからの市況分析コメントの自動生成人工知能学会研究会 SIG-FIN-020
- 3) T. Luong, H. Pham, and C. D. Manning. Effective approaches to attentionbased neural machine translation. In EMNLP, 2015
- 4) Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111- 3119, 2013.
- 5) Kudo, T.: Mecab: Yet another part-of-speech and morphological analyzer, <http://mecab.googlecode.com/>(2005)

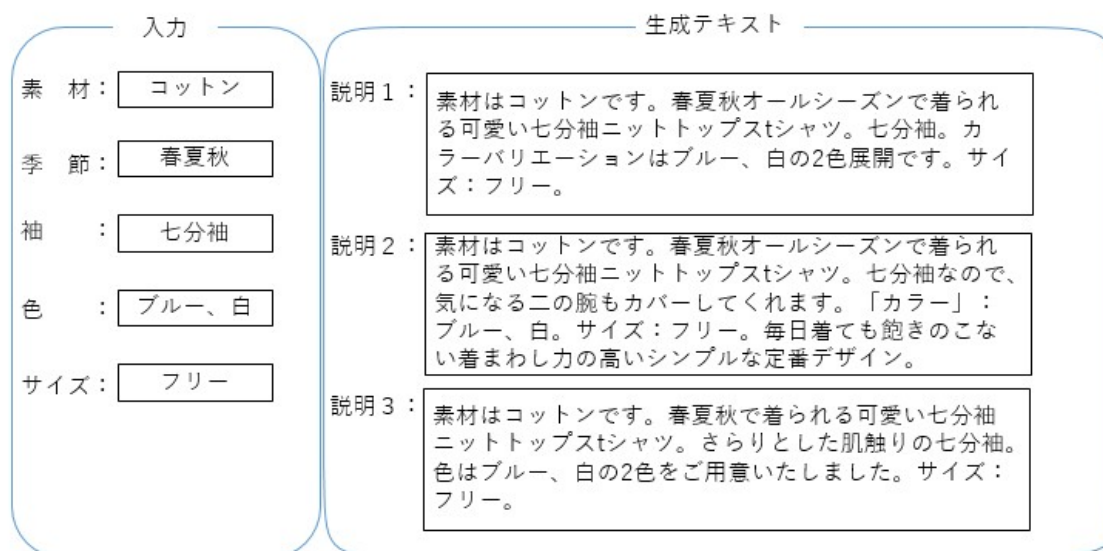


図2 生成モデルのサンプル

カテゴリ	商品	商品特徴					言葉通じ
		素材	季節	袖など	カラー	サイズ	
レディースTシャツ	商品 1	1	1	1	1	1	0.5
	商品 2	0.6	1	0.4	1	1	0.4
	商品 3	1	1	1	1	1	0.5
	商品 4	1	0.7	0.2	1	1	0.4
	商品 5	1	0.6	0.9	1	0.9	0
	商品 6	1	0.9	0.6	1	0.9	0.2
	商品 7	1	0.7	0.9	1	1	0.7
	商品 8	0	1	1	1	1	0
	商品 9	1	0.7	0.8	1	1	0.2
	商品 1 0	0	0.7	0.7	1	0.8	0.3
	平均	0.76	0.82	0.73	1	0.96	0.34
レディースブラウス	商品 1	1	1	1	0	1	0.6
	商品 2	1	1	1	1	1	0.4
	商品 3	1	0.9	0.8	1	1	0.5
	商品 4	1	1	1	1	1	0
	商品 5	1	1	1	1	1	1
	商品 6	1	0.8	1	1	1	0.2
	商品 7	1	1	0.9	0.8	1	0.3
	商品 8	1	1	1	1	1	0
	商品 9	1	1	0.9	1	1	0.9
	商品 1 0	1	0.8	1	0.6	1	0.7
	平均	1	0.95	0.96	0.84	1	0.46
メンズトレーナー	商品 1	1	1	0.9	0.8	1	0.3
	商品 2	1	0.9	1	0.8	1	1
	商品 3	0.9	1	1	1	1	0.8
	商品 4	1	1	1	0.6	1	0.3
	商品 5	1	1	1	0.9	1	0.6
	商品 6	1	1	1	1	1	0.6
	商品 7	1	1	1	1	1	0.9
	商品 8	1	1	1	0.8	1	0.4
	商品 9	1	1	1	1	1	0.9
	商品 1 0	1	1	1	1	1	0.4
	平均	0.99	0.99	0.99	0.89	1	0.62

表3 テキスト生成評価