

PersuAIDE ! An Adaptive Persuasive Text Generation System for Fashion Domain

Vitobha Munigala
IBM Research
vmunigal@in.ibm.com

Abhijit Mishra
IBM Research
abhijimi@in.ibm.com

Srikanth G Tamilselvam
IBM Research
srikanth.tamilselvam@in.ibm.com

Shreya Khare
IBM Research
shkhare4@in.ibm.com

Riddhiman Dasgupta
IBM Research
riddasgu@in.ibm.com

Anush Sankaran
IBM Research
anussank@in.ibm.com

ABSTRACT

Persuasiveness is a creative art which aims at inducing certain set of beliefs in the target audience. In an e-commerce setting, for a newly launched product, persuasive descriptions are often composed to motivate an online buyer towards a successful purchase. Such descriptions can be catchy taglines, product-summaries, style-tips *etc.*. In this paper, we present *PersuAIDE!* - a persuasive system based on linguistic creativity to generate various forms of persuasive sentences from the input product specification. To demonstrate the effectiveness of the proposed system, we have applied the technology to fashion domain, where, for a given fashion product like “*red collar shirt*” we were able to generate descriptive sentences that not only explain the item but also garner positive attention, making it persuasive.

PersuAIDE! identifies fashion related keywords from input specifications and intelligently expands the keywords to creative phrases. Once such compatible phrases are obtained, persuasive descriptions are synthesized from the set of phrases and input keywords with the help of a neural language model trained on a large domain-specific fashion corpus. We evaluate the system on a large fashion corpus collected from different sources using (a) automatic text generation metrics used for Machine Translation and Automatic Summarization evaluation and Readability measurement, and (b) human judgment scores evaluating the persuasiveness and fluency of the generated text. Experimental results and qualitative analysis show that an unsupervised system like ours can produce more creative and better constructed persuasive output than supervised generative counterparts based on neural sequence-to-sequence models and statistical machine translation.

KEYWORDS

Persuasiveness, Persuasive Systems, Persuasion in Fashion, Style-tip Generation

ACM Reference format:

Vitobha Munigala, Abhijit Mishra, Srikanth G Tamilselvam, Shreya Khare, Riddhiman Dasgupta, and Anush Sankaran. 2018. PersuAIDE ! An Adaptive

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WWW’18, Lyon France

© 2018 ACM. 978-x-xxxx-xxxx-x/YY/MM...\$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

Persuasive Text Generation System for Fashion Domain. In *Proceedings of ACM The web conference, Lyon France, April 2018 (WWW’18)*, 8 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

One of the major drawbacks of the current *Weak AI* based language generation systems is the inability to generate language in the same manner, expressing similar levels of creativity, originality and brevity as humans. To produce language generators that are more humanly (otherwise known as Strong AI systems [17]), one has to demonstrate the capabilities of such systems towards performing generation tasks that demand extreme cognitive capabilities. In this paper, we proceed in such a direction, with the aim of empowering natural language generation systems with the ability to be *persuasive*. Persuasiveness, which is an extremely creative and cogitative art, often intends to infuse cognitive change in the mental state of target audience. Persuasive systems have utilities in multitude of scenarios - dynamic advertisement, preventive medicine, social action, and edutainment to name a few, as mentioned in [11].

We present *PersuAIDE!* - a persuasive system that brings in linguistic creativity to transform given specification (a list of keywords) to various forms of persuasive descriptions; the descriptions are typically one sentence long. To demonstrate the effectiveness, we apply the technology to fashion domain, where, for given keywords of a fashion item like *Levis white round neck t-shirt*, the system generates descriptive sentences that not only explains the item but also garners positive attention, making it persuasive (e.g., “*Accentuate an easy-going summer look with white round neck t-shirt¹*”). Our system can be quite useful especially in an fashion e-commerce setup, as illustrated in Figure 1, which is from a popular fashion e-commerce website of TATACLIQ².

Persuasive systems are nuanced by two key challenges as mentioned below.

- (1) **Data scarcity:** It is almost impossible to generate domain specific labeled corpus (parallel corpus) containing general descriptions and the corresponding persuasive descriptions. This is due to the fact that there can be a considerably large number of possible persuasive variations of the given input descriptions. For the input of “*Levis white turtle neck t-shirt*”, several persuasive descriptions can be generated, emphasizing on each individual aspect such as “*Levis*”, “*White*”, “*Turtle neck*” and “*T-shirt*”. Hence, it is hard to

¹ This description has been generated by our system

² <https://www.tatacliq.com/>

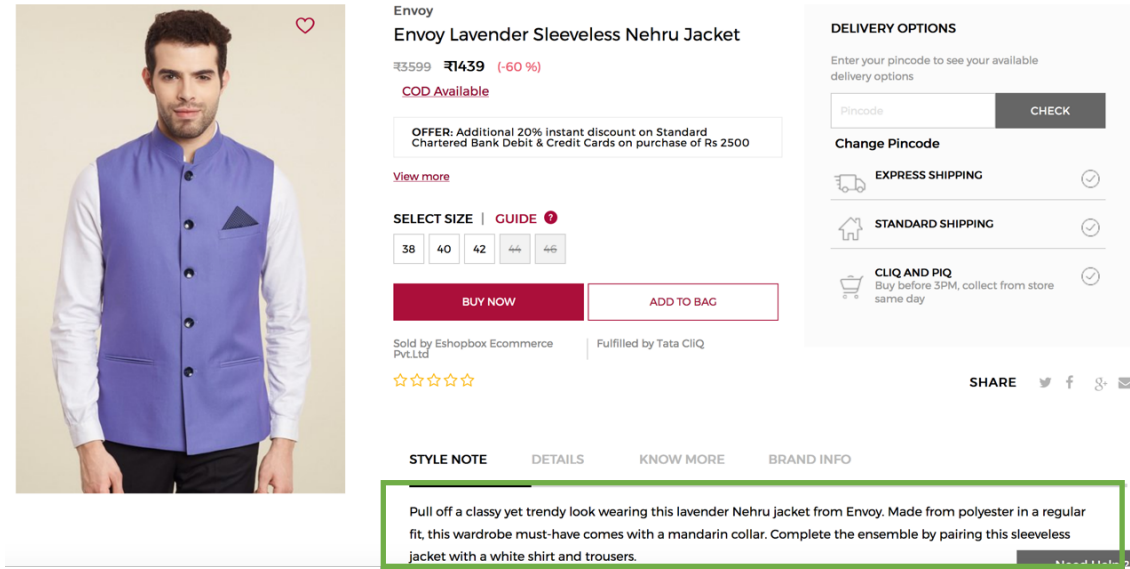


Figure 1: Illustrative example from TataCliq

run popular supervised machine learning / deep learning based transformation techniques such as Statistical Machine Translation [15] or Neural Sequence-to-sequence Transformation [28] or their variants.

- (2) **Lack of good evaluation measures:** Evaluating the persuasiveness of a given text can be highly challenging as it requires jointly evaluating the *psychological and cognitive* aspects of persuasiveness (such as creativity, originality) as well as the linguistic aspects (such as fluency and adequacy) of the text generated.

PersuAIDE! intends to tackle the first challenge by following a modular architecture, in which the modules, at most, rely on a large amount of monolingual, in-domain English corpora and word list that are easily accessible. The job of transforming an input specification into a persuasive description is carried out by (a) parsing the input specification to obtain a list of entities, (b) deciding key-entities on which the text has to be persuasive, (c) mining a list of relevant persuasive key-phrases which could be suitable to describe the key-entities, and (d) stitching the persuasive key-phrases and entities in the input specification together to generate the persuasive description. The systems performance is evaluated using various manual and automatic evaluation measures which test the system from psychological, cognitive and linguistic perspectives. Experiments indicate that our system, which does not require any labeled data for persuasive text generation, performs competitive to supervised machine learning based alternatives in the lines of persuasiveness and fluency.

The rest of the paper is organized as follows. Section 2 presents a summary of works that are relevant to ours. Section 3 introduces the overall system architecture and the components. The experiment setup is discussed in Section 4. Evaluation results and error-analysis are presented in Section 5. Section 6 concludes the paper and points to future possibilities.

2 RELATED WORK

Research in creative natural language text generation has seen a spurt in recent years which is evident from the classes of works on poetry generation[9][29][5][24][8], story generation[13][22], slogan generation[30] [33], humor generation [32] [31]. Similarly, there has been a good amount of active study on persuasion system and its various aspects[27][6][10]. In this section, we provide an overview of two work that are most relevant and heavily inspired the development of *PersuAIDE!*.

Ozbal et al. [25] proposed an extensible framework BRAINSUP that allows retaining of input key-words in the final expression, inducing user-desired emotion, generating domain specific expressions *etc.* But it makes heavy use of syntactic information (in other words, templates) to generate well-formed sentences and generation happens by selecting an appropriate template and then filling the empty slots in the template with appropriate words from the input.

Lexical substitution based methods have been used for generating persuasive sentences. Tomasic et al. [30] follow the same approach as BRAINSUP to transform slogan to target specification. Gatti et al. substitute popular concept from trending news in template expressions [7]. But these works consume fairly expressive text like headline, slogan as inputs and apply simple substitution for domain adaptation where as *PersuAIDE!* generates complete sentences from roughly 3 to 4 keywords.

A notable study focusing on negotiation, a close form to persuasion, Facebook Artificial Intelligence Research (FAIR) team's work on *Deal or no deal?: Training AI bots to negotiate* [18] demonstrate linguistic and reasoning capabilities of the AI bots to negotiate. While their work is a pioneering work in negotiations, their utterances are limited and are focused towards reaching consensus rather than persuading the other agent based on catchy expressions. Moreover, since it is fundamentally a dialog system, it gets

Abbreviation	Description
DC	Domain corpus
D	Product Descriptions
DNP	Domain Noun Phrase
B	List of brands
P	List of product categories
C	List of colors
DA	List of domain adjectives
DV	List of domain verbs

Table 1: List of abbreviations and descriptions

several chances during a conversation to understand the opponent agent and act accordingly. A system like *PersuAIDE!* gets only one chance to appeal to the global consumers with no prior knowledge regarding possible reactions from users.

3 SYSTEM ARCHITECTURE

Our system takes a short product specification text as input (henceforth referred to as “*title text*”) and produces a persuasive description as the output (henceforth referred to as “*style tip*”). Figure 2 shows an overview of the proposed system architecture with a suitable end-to-end example. The proposed system has five major steps: (i) Given the title text of a product, the first step is identify the entities and tag them as generic keywords or named entities (such as brand names), (ii) a keyword expansion is then performed to enhance the generic keywords that encompass more qualitative information about the product, (iii) a set of relevant catchy noun phrases are selected based on the enhanced keywords; these phrases are supposed to best explain the given title text, (iv) using a neural language model, the selected noun phrases are combined with the extracted entities and a prior list of function words to yield candidate persuasive sentences, (v) ranking is performed on the candidate sentences generated in the previous step to choose the most persuasive and fluent style tip for the given title text.

In further sections, we provide a detailed explanation of each module in our system.

3.1 Resources Used

Prior to explaining the system components, we describe the data resources used to implement the different modules of our system. We would like to remind the reader that our system takes no labeled data as input, *i.e.*, it does not require any parallel corpora containing “title-text” at one side and “style-tip” at the other side. This helps us overcome the first challenge of persuasive text generation *i.e.*, data-scarcity, as mentioned in the introductory section. The resources we intend to use in our system are either in the form of lexicons (a list of domain specific relevant words) or a large amount of unlabeled domain specific texts.

Table 1 captures notations used in our work. We now list the resources used below.

3.1.1 Domain Corpus. We extract an unlabeled corpus related to fashion domain (referred to as *DC*) from various retail sites and magazines. The texts present fashion products from various categories. For each product, we extract the product title, product

description or the style-tip, product brand, fabric, suitable gender, etc. Overall, we collected around 149k sentences related to product descriptions. From some websites, we could extract both “title-text” and “style-tip” for the same product, which resulted in a small scale parallel-corpus. We do not use the “title-text” and “style-tip” alignment anywhere in our pipeline and whatever parallel corpus is extracted is used to train and test out supervised baselines such as sequence-to-sequence model and statistical machine translation.

3.1.2 Domain constructs. We populate useful fashion constructs namely adjectives (*DA*), verbs (*DV*) from *words-to-use* website³. Additionally, a list of brand names (*B*), product types (*P*), and colors (*C*) are extracted from the e-commerce website data discussed above using simple heuristics.

3.1.3 Domain Noun phrases (DNP). From the crawled unlabeled fashion corpus *DC*, product descriptions *D* are extracted, we bootstrap useful descriptive noun phrases *DNP* like *easy-going summer look*, *real heart stealer*, *a tassel-tipped tie-up*. These noun phrases are extracted using the following steps:

- **Step 1:** Initialize *DNP* list with NULL.
- **Step 2:** Sentence-tokenize *D*.
- **Step 3:** Perform noun chunking of each sentence *S* in *D*. This will extract a list of noun phrases (*NP*) from *S*.
- **Step 4:** If *NP* satisfies constraints 1 and 2 (mentioned below), add it to *DNP* list.

For tokenizing and noun-chunking we used the *Spacy* parser. The noun-phrases in *DNP* are expected to be generic descriptors and should not represent a specific brand, color, or product. To ensure this, two constraints are imposed: (1) the noun phrase should be between 3-5 words and (2) the noun phrase should not contain words from *B*, *P* and *C*, which are highly product/brand/item/color specific.

3.1.4 Domain Word Representations. In order to build a meaningful context-aware representation for each word in the fashion domain corpus, we trained word2vec[23] model using the title text and product description sentences in *DC*. This word2vec model has been exclusively trained on a fashion corpus to capture the semantic relationship between the words in the fashion domain.

3.2 Input Preprocessing and Entity Separation

The input title text is parsed to identify the keywords. This lists all the words in the input title text; an example is shown in Figure 2. The keywords are then categorized into two classes: (1) **Generic Keywords:** words that carry information about color, gender, and product type and (2) **Named Entities:** words containing brand names, named products, for example, “Nehru” in “Nehru Jacket”. Generic keywords are then sent to the keyword expansion module whereas the named entities are retained to be used in the sentence generation module.

3.3 Keywords Extraction and Expansion

This module expands the generic keyword list by adding more nouns, adjectives, and adverbs to it. The idea here is to get more

³<http://www.words-to-use.com/words/clothing/>

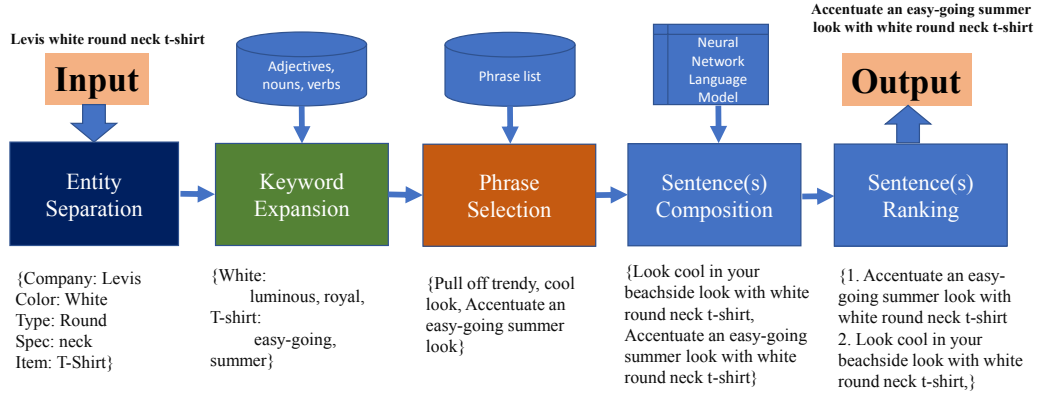


Figure 2: Architecture for persuasive text generation

informative and creative words which, in some sense, are closely associated with the generic keywords. For our purpose, a word-based similarity measure is implemented using a similarity measure between two word-embedding-vectors (like cosine similarity) is thus used to measure the semantic relatedness among the words. For the input keyword i , the semantic relatedness score is used to pick top K similar words from the fashion vocabulary. To eliminate possibilities of adding noisy/irrelevant words (e.g., “brown” could turn out to be highly similar to “white” in the embedding space but it is not a good candidate in our case), we impose additional constraints such as the part-of-speech of the related keywords should be either “common noun” or “adjective” and should not match with words from B , P and C , which are highly product/brand/item/color specific.

To give an example, for the input title text *levis white neck t-shirt*, the term ‘levis’ is identified as a brand and is therefore ignored. The rest of the terms are expanded to obtain words like ‘luminous’, ‘royal’, ‘retro’, ‘funky’ etc based on *word2vec* similarity and constraint satisfaction. At this point we have K descriptors for each generic keyword entity.

3.4 Noun phrase selection

The K expanded keywords are then compared with the domain noun phrase list DNP and the top L phrases are identified, again based on the maximum similarity score. At this point, we are aware of the best noun phrases that goes well with expanded keywords. Example ‘easy-going summer look’ might be the best noun phrase for ‘retro t-shirt’. All the unique DNP are selected for sentence generation task.

Since the comparison of the expanded keywords and the noun phrases happen in the *word2vec* space, it is important for us to get a common representation for both the keywords as well as the phrases. For keyword representation, we take the average between the word embedding for generic keyword k and the word embedding of the expanded keywords. For DNP representation, the embeddings of the constituent words are averaged. Once the representations for both the entities are obtained, a semantic relatedness score based on *Earth Mover’s distance (EMD)* is established

between all the possible pairs of feature modifiers, DNP . Based on the matrix of semantic relation scores; the best score for each DNP is obtained and finally, the top L from DNP list are chosen to be part of the sentence composition.

3.5 Sentence Generation

The keywords extracted from the input title text, top L domain noun phrase from DNP list, common function/stop words S , domain verb list DV , and a stop word list are used to prepare the final sentence. We leverage domain adjectives DA , only to identify the compatible noun phrases for the input keywords and they are not used in generation task. By this step, relevant persuasiveness is induced in the form of expanded keywords and domain noun phrases. Further, persuasiveness is induced with the introduction of verbs that determines how the candidates could be used to form a fluent (and hopefully, catchy) sentence. We leverage neural language model [2] for this final task of sentence generation. During generation, *beam search* is applied to prevent the explosion of possible hypotheses (pertaining to all possible choices from candidates words and phrases) at a given time step in the generation. Our generation ensures that selection of next words/phrases in each time step minimizes the overall perplexity of the description generated so far with respect to the language model.

We initiate the generation process with the selection of a seed word (which is always a verb in our case as we aim to generated imperative persuasive texts). The seed word is randomly sampled from the DA . At each time step a word/phrase is chosen in a way such that the overall perplexity score is minimized. The generation aims to consume all the keywords extracted from the title text, L DNP with an appropriate usage of the function words to coherently stitch the essential components of the sentence.

A concise representation of the entire system is provided in Algorithm 1.

Algorithm 1 Style-Tip Generation

```

1:  $LM \leftarrow$  Load the Neural Language Model
2:  $SimModel \leftarrow$  Load the word2vec model trained on the fashion corpus
3:  $fashionVerbs \leftarrow$  Load the fashion related verbs obtained from external source
4:
5: procedure GENERATE( $input, brandName, beamSize$ )
6:    $title \leftarrow$  remove  $brandName$  from  $input$ 
7:    $keywordsList \leftarrow$  extract keywords from the  $input$ 
8:   for keyword  $K$  in  $keywordsList$  do
9:      $extendedWords \leftarrow$  keyword expansion for  $K$  using  $simModel$ 
10:     $relatedPhrases \leftarrow$  select related phrases from the list of fashion phrases using the keyword  $K$  and  $extendedWords$ 
11:     $keyList \leftarrow relatedPhrases + rootWord$ 
12:     $stopWords \leftarrow$  List of prepositions
13:     $seedWord \leftarrow$  sample word from the list of  $fashionVerbs$ 
14:     $generations \leftarrow$  Initiate beam search BEAM( $seedWord, keyList, stopWords, beamSize$ ) to obtain the best generations starting with  $seedWord$  and making use of  $keyList$  and  $stopWords$ 
15:
16: procedure BEAM( $currentGeneration, keyList, stopWords, beamSize$ )
17:   while no new word to add do
18:      $generation \leftarrow$  add best word to the  $currentGeneration$  from  $keyList + stopWords$  by making use of the  $LM$ 
19:      $currentGeneration \leftarrow$  select  $beamSize$  generations by making use of  $LM$  score for each generation
20: return  $currentGeneration$ 

```

4 EXPERIMENTAL SETUP

The dataset and the setup used in this research to experimentally validate the proposed persuasive system is described in this section. Further, the evaluation metrics are defined and some existing systems are explained that are used for comparison purposes.

4.1 Dataset

As discussed earlier, we design the system to work on fashion domain data crawled from various websites. Some statistics about the dataset are mentioned in the Table 2. Cumulatively, our dataset consists information about 149K products spanning different categories like men’s wear, women’s wear, footwear, etc. For each product, we extract multiple parameters like title, description, color, gender, age group, taxonomy, brand name, etc. However, for the current task, we mainly rely on the product title and its description.

4.1.1 Dataset Split for Evaluation and Comparison. As we have pointed in the introductory section, it is difficult to apply popular supervised generation techniques for persuasive text generation (such as Sequence to Sequence neural models, Statistical Machine Translation etc.) as getting parallel labeled text would be extremely

	value
Number of samples	149320
Avg. words in title	6
Max. words in title	15
Avg. words in style-tip	18
Max. words in style-tip	74
Number of colors	696
Number of brands	591
Total samples related to men	95677
Total samples related to women	53559
Number of fashion categories	234

Table 2: Dataset Statistics

difficult. We still managed to create a small parallel corpus from a retail website. The title text is considered as the source side text and the style tip is considered as the target text. This way we could gather a parallel corpus which is split to a train-test set of in the ratio of 93:7.

The performance of our system is compared with some existing systems using (discussed in 4.2) using known evaluation metrics (discussed in 4.3), using the test data of 2736 instances. Additionally, we perform manual evaluation by randomly forming a held-out set of 30 test-instances.

4.2 Comparison with Existing Systems

Since there is no other existing system at this moment that works with a similar goal like ours, we decide to compare our system’s performance with three relatively close systems. The systems are mentioned below.

- (1) **Slogan4U:** slogan4u.com is a free slogan generator that takes a phrase as input and produce slogan variations as output. The reason we chose this as a viable system for comparison is that the slogan generated by the website (a) retain the input phrase in them, thereby offering a loss-less transformation of the input phrase (b) the slogans appear to be creative and, to some extent, persuasive.
- (2) **Seq2Seq:** We train a neural sequence-to-sequence system [20] built with Long-Short-Term-Memory (LSTMs) units over input word embeddings. The framework is based on encoder-decoder architecture and is empowered with attention mechanism. We use default parameters and training configuration as reported by the authors of [20]. The validation set is used to tune other hyperparameters.
- (3) **SMT:** We train a *phrase-based statistical machine translation* (SMT) system using the Moses framework [16]. For phrase based translation model learning, *grow-diag-final-and* heuristic is opted, and to tackle lexicalized reordering, we use the *msd-bidirectional-fe* model. GIZA++ is configured to apply the principles of IBM model 4 and 5 for alignment learning [3]. We tuned the trained SMT models with the validation set using Minimum Error Rate Training (MERT) with default parameters (100 best list, max 25 iterations). We trained a 5-gram language model using the target side (descriptions) of the training data. For this, we

System	BLEU-4 (%)	METEOR	ROUGE 1	ROUGE L	Similarity
Solgan4U	19.6	0.16	0.14	0.11	0.76
Seq2Seq	25.91	0.21	0.26	0.20	0.64
SMT	56.91	0.24	0.22	0.2	0.7
Our Approach @1	22.0	0.23	0.17	0.21	0.75
Our Approach @2	22.4	0.24	0.18	0.22	0.74

Table 3: Evaluation of systems using metrics used for Machine Translation and Summarization.

System	Flesch-Kincaid	SMOG-Index	DaleChal-Index
Reference	7.38	8.81	9.85
Solgan4U	4.4	6.3	10.7
Seq2Seq	5.68	5.9	10.6
SMT	4.05	6.06	9.38
Our Approach @1	9.02	9.76	12.18
Our Approach @2	8.61	9.31	11.61

Table 4: Evaluation of systems using different readability measures (averaged across test data)

System	Catchyness	Relatedness	Fluency
slogan4u	0.4977	0.6244	0.68
slogan4u	0.4911	0.6755	0.7533
Our Approach @1	0.7688	0.9	0.7688
Our Approach @2	0.7	0.9022	0.7

Table 5: Manual evaluation of systems using different metrics for persuasiveness

use use Kneser-Ney smoothing algorithm implemented in KenLM toolkit [12], bundled with Moses.

4.3 Evaluation

Evaluating a creative task like persuasive text generation is challenging as it requires evaluating the system from both psychological and linguistic perspectives. In the absence of direct evaluation metrics, we choose the following automatic evaluation metrics that are popularly used for generation tasks (especially Machine Translation and Summarization). Additionally, we use popular readability metrics to check the quality of the generated descriptions from the perspective of readability.

The evaluation metrics are listed below:

(1) **Generation:**

- **BLEU:** A popular n-gram match based evaluation metrics [26].
- **METEOR:** Based on n-gram match but also considers synonyms and paraphrase based replacements [1].
- **ROUGE:** A popular *recall-based* metric used for summarization evaluation [19].
- **Similarity:** Document similarity method using *Spacy* tool indicating to what extent the generated text is similar to the input title text.

(2) **Readability:**⁴

- **Flesch-Kincaid Index** [14]

⁴Higher readability scores indicate more reading complexity

- **SMOG Index** [21]
- **Dale-Chal Index** [4]

Since, the above metrics are still not enough to capture the creative aspects of persuasive text generation, we carry out manual evaluation on a small test dataset of 30 instances. The metrics that we use for manual evaluation following the work of [25], are as follows:

- (1) **Catchyness:** Is the description attractive, catchy [Yes/No]?
- (2) **Relatedness:** Is the description semantically related to the target domain? [Yes/No]?
- (3) **Fluency:** Is the sentence grammatically correct? [Yes/No]?

We now proceed to the result section.

5 RESULTS

For our system, we take two different output candidate descriptions obtained with respect to two different sets of expanded keywords from the input title-text. The two output descriptions are denoted as @1, @2 approach in result tables. Results are reported on the test data of 2736 instances, used for automatic evaluation and a held-out set of 30 instances, used for manual evaluation.

From the perspective of considering our persuasive system as yet another machine translation system that converts a non-persuasive product title into a persuasive product description, we have results shown in Table 3. Popular machine translation and summarization metrics such as BLEU, METEOR, ROUGE-1 and ROUGE-L, and similarity measures are used. The proposed algorithm is compared against three existing systems. Since most of these metrics work purely based on the keyword matching, most of the creative systems tend to perform low because of the artistic freedom to choose varied and unusual words. Additionally, methods like sequence-to-sequence and SMT learn to produce results closer to the reference text, indirectly optimizing the evaluation metrics we have considered here, hence the results are better for these techniques. But, given the fact that it is extremely difficult to get access to parallel data in real world scenarios, our system, producing results competitive to these supervised systems is definitely a positive indication going forward.

From the perspective of generative systems, we use several readability measures to evaluate if the persuasive generated descriptions are readable (in terms of English language structure). The results are shown in Table 4. Across all the three measures: Flesch-Kincaid, SMOG-Index, and DaleChal-Index, we observe that our proposed system generates quite readable sentences (with scores ranging from [9-12]). However, the scores being higher for our systems than the baselines indicated that descriptions generated by our systems contain more nuances (from the perspective of reading) than the simpler descriptions generated by the baselines.

From the perspective of human psychology of persuasive product descriptions, we manually evaluated the generated descriptions using human interviews. Three different measures were used to evaluate the human subjectiveness: Catchyness, Relatedness, and Fluency. It can be evidently observed that the proposed system generated more catchy, better related, and grammatically correct sentences compared to the creative slogan generation systems.

Input	oxoloxo orange checks shirt
Reference	Opt for casual comfort with this orange shirt from oxoloxo
slogan4u	oxoloxo orange checks shirt the cream of the crop.
seq2seq	Look stylish by sporting this pair of oxoloxo casual shoes.
SMT	This orange color shirt from oxoloxo
PersuAIDE!	Go well with your off-duty styles on a stylised fit using orange checks shirt
Input	cobblerz black derby formal shoes
Reference	Embrace the age old charm of derby shoes with the black derby shoes by cobblerz.
slogan4u	Got cobblerz black derby formal shoes?
seq2seq	Flaunt your casual collection with this black top from cobblerz. Made from polyester fabric, this regular fit top will keep you sweat free all day long.
SMT	Flaunt your chic look with black shoes
PersuAIDE!	Embellish at a chic finishing touch with the new season's trends using black derby formal shoes
Input	zudio dark beige back strap sandals
Reference	Update your footwear collection with 's dark beige sandals.
slogan4u	zudio dark beige back strap sandals is crazy good.
seq2seq	Team this shirt with a pair of denims and casual shoes.
SMT	fetherlite this pair of dark brown casual
PersuAIDE!	Flaunt on your amazing fashion sense in a swanky ensemble using dark beige back strap sandals.
Input	109 f black printed cavi top
Reference	Incorporate unique prints in your wardrobe with this black Cavi top from 109f.
slogan4u	109 f black printed cavi top quality you can see.
seq2seq	Look stylish by wearing this pair of 109f casual shoes.
SMT	Steal the limelight wearing this pair of casual shoes from
PersuAIDE!	Flatter the figure this black printed cavi top with a marvelously versatile pick using flavoursome fusion fashion
Input	dc council mid lx navy sneakers
Reference	Hit the road in style by wearing this pair of navy casual sneakers from the house of dc council
slogan4u	Me and my dc council mid lx navy sneakers.
seq2seq	Look stylish by sporting this pair of dc council casual shoes.
SMT	sultan in style by wearing this pair of navy casual sneakers from
PersuAIDE!	Look sexy in your beachside look with mid lx navy sneakers using a cool hip-hop look

Table 6: Sample generations from different systems along with inputs and reference descriptions

It can be clearly observed from Table 3 and Table 5 that, though our system outperforms the existing system based on human evaluation, it performs low on the standard text generation metrics. This can be attributed to the constraint of matching the exact words or synonyms and not considering the creative liberty to choose words/phrases that are unusual.

For qualitative analysis, we also provide the sentences generated from our system as well as other systems in the Table 6. As we can see, the descriptions generated by our systems are competitive or better in terms of creativity, persuasiveness and fluency than the supervised baselines but have less overlap with the reference descriptions. This explains why our system is deemed to have underperformed than the baselines, as per the automatic evaluation scores. In general, the field of creative text generation demands looking beyond simplistic evaluation measures and it is about time that trainable metrics for evaluating persuasive text holistically, including aspects on creativity, coherency, novelty are proposed.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented a system called PersuAIDE! which aims to generate persuasive text for input product description. We demonstrated the work in fashion domain where for a given fashion product title such as *'Levis white round neck t-shirt'*, we were able to first expand features and keywords of the provided entity words in the title. Further, we generate descriptive sentences with random seed verbs using a neural network based language model and other NLP techniques to generate a persuasive sentence - *shape your evening outfit with this white neck t-shirt*. We evaluate our work against a popular slogan generation system using manual

analysis and compare our work with popular supervised generative models. We also find that our work consistently performs as good as supervised machine learning based alternatives in the lines of persuasion and correctness, as established through various metrics. Our study could be considered as a novel attempt to generate persuasive product description in a linguistically motivated framework that accounts for syntagmatic and paradigmatic aspects of language. Applying our technique to different other domains (such as news, tourism *etc.*) is definitely a future possibility.

REFERENCES

- [1] Satandeep Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, Vol. 29. 65–72.
- [2] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. 2003. A neural probabilistic language model. *Journal of machine learning research* 3, Feb (2003), 1137–1155.
- [3] Peter F Brown, Vincent J Della Pietra, Stephen A Della Pietra, and Robert L Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational linguistics* (1993).
- [4] Jeanne Sternlicht Chall and Edgar Dale. 1995. *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.
- [5] Simon Colton, Jacob Goodwin, and Tony Veale. 2012. Full-FACE Poetry Generation.. In *ICCC*. 95–102.
- [6] Lorenzo Gatti, Marco Guerini, Oliviero Stock, and Carlo Strapparava. 2014. Sentiment variations in text for persuasion technology. In *International Conference on Persuasive Technology*. Springer, 106–117.
- [7] Lorenzo Gatti, Gozde Ozbal, Marco Guerini, Oliviero Stock, and Carlo Strapparava. 2016. Hedy-Lines: A Creative Generator Of Newspaper Headlines. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces*. ACM, 79–83.
- [8] Hugo Gonalo Oliveira. 2015. Tra-la-lyrics 2.0: Automatic generation of song lyrics on a semantic domain. *Journal of Artificial General Intelligence* 6, 1 (2015), 87–110.

- [9] Erica Greene, Tugba Bodrumlu, and Kevin Knight. 2010. Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 524–533.
- [10] Marco Guerini, Fabio Pianesi, and Oliviero Stock. 2014. Is it morally acceptable for a system to lie to persuade me? *arXiv preprint arXiv:1404.3959* (2014).
- [11] Marco Guerini, Oliviero Stock, and Massimo Zancanaro. 2007. A taxonomy of strategies for multimodal persuasive message generation. *Applied Artificial Intelligence* 21, 2 (2007), 99–136.
- [12] Kenneth Heafield. 2011. KenLM: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*.
- [13] Parag Jain, Priyanka Agrawal, Abhijit Mishra, Mohak Sukhwani, Anirban Laha, and Karthik Sankaranarayanan. 2017. Story Generation from Sequence of Independent Short Descriptions. *arXiv preprint arXiv:1707.05501* (2017).
- [14] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Technical Report. DTIC Document.
- [15] Philipp Koehn. 2009. *Statistical machine translation*. Cambridge University Press.
- [16] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*. Association for Computational Linguistics, 177–180.
- [17] Ray Kurzweil. 2005. *The singularity is near: When humans transcend biology*. Penguin.
- [18] Mike Lewis, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or No Deal? End-to-End Learning for Negotiation Dialogues. *CoRR abs/1706.05125* (2017). [arXiv:1706.05125](http://arxiv.org/abs/1706.05125) <http://arxiv.org/abs/1706.05125>
- [19] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop*, Vol. 8. Barcelona, Spain.
- [20] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025* (2015).
- [21] G Harry Mc Laughlin. 1969. SMOG grading-a new readability formula. *Journal of reading* 12, 8 (1969), 639–646.
- [22] Neil McIntyre and Mirella Lapata. 2009. Learning to tell tales: A data-driven approach to story generation. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*. Association for Computational Linguistics, 217–225.
- [23] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
- [24] Hugo Gonçalves Oliveira. 2012. PoeTryMe: a versatile platform for poetry generation. *Computational Creativity, Concept Invention, and General Intelligence 1* (2012), 21.
- [25] Gözde Özbal, Daniele Pighin, and Carlo Strapparava. 2013. BRAINSUP: Brainstorming Support for Creative Sentence Generation.. In *ACL (1)*. 1446–1455.
- [26] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 311–318.
- [27] Oliviero Stock, Marco Guerini, and Fabio Pianesi. 2016. Ethical Dilemmas for Adaptive Persuasion Systems.. In *AAAI*. 4157–4162.
- [28] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.
- [29] Jukka Toivanen, Hannu Toivonen, Alessandro Valitutti, Oskar Gross, et al. 2012. Corpus-based generation of content and form in poetry. In *Proceedings of the Third International Conference on Computational Creativity*.
- [30] Polona Tomašić, M Znidaršić, and Gregor Papa. 2014. Implementation of a slogan generator. In *Proceedings of 5th International Conference on Computational Creativity, Ljubljana, Slovenia*, Vol. 301. 340–343.
- [31] Alessandro Valitutti, Oliviero Stock, and Carlo Strapparava. 2009. Graphlaugh: a tool for the interactive generation of humorous puns. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 1–2.
- [32] Alessandro Valitutti, Hannu Toivonen, Antoine Doucet, and Jukka M Toivanen. 2013. "Let Everything Turn Well in Your Wife": Generation of Adult Humor Using Lexical Constraints.. In *ACL (2)*. 243–248.
- [33] Martin Znidaršić, Polona Tomašić, and Gregor Papa. 2015. Case-Based Slogan Production. (2015).