COMMA: Modeling Relationship among Motivations, Emotions and Actions in Language-based Human Activities

Yuqiang Xie Yue Hu[†] Wei Peng Guanqun Bi Luxi Xing
Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
{xieyuqiang,huyue,pengwei,biguanqun,xingluxi}@iie.ac.cn

Abstract

Motivations, emotions, and actions are inter-related essential factors in human activities. While motivations and emotions have long been considered at the core of exploring how people take actions in human activities, there has been relatively little research supporting analyzing the relationship between human mental states and actions. We present the first study that investigates the viability of modeling motivations, emotions, and actions in language-based human activities, named **COMMA** (**Co**gnitive Fra**m**ework of Hu**m**an **A**ctivities). Guided by **COMMA**, we define three natural language processing tasks (emotion understanding, motivation understanding and conditioned action generation), and build a challenging dataset **HAIL**[‡] through automatically extracting samples from Story Commonsense. Experimental results on NLP applications prove the effectiveness of modeling the relationship. Furthermore, our models inspired by **COMMA** can better reveal the essential relationship among motivations, emotions and actions than existing methods.

1 Introduction

Human activities are continuous interactions between external environment (physical world and social events, etc.) and internal mind (motivations, emotions, etc.). For example, Figure 1 demonstrates human activities of character 'I' about 'eating bread' in external environment, as well as the mental states of 'I'. The motivation of character 'I' is a **physiological** need. Conditioned on this motivation and history actions, 'I' **enjoyed the part of bread** on the direction of **joy** emotion. While human mental states have long been considered at the core of exploring how people take actions between the lines in language-based human activities, there has been relatively little research supporting analyzing the relationship between human mental states and actions. It is challenging to comprehensively model the relationship of motivations, emotions and actions in language-based human activities, which can allow researchers to reason the essential causes of human activities from the cognitive perspective and supply reasonable explanations. This technology will have a profound impact on various natural language processing (NLP) downstream applications, such as intelligent dialogue, controllable text generation, recommendation systems, and public opinion analysis.

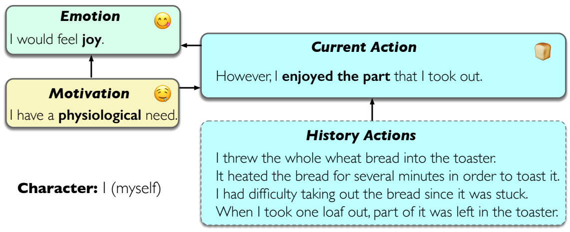


Figure 1: An example of human activity. History actions and motivation cause current action. And emotion is the effect of motivation and actions. Here, motivation/action/emotion is colored by yellow/cyan/green.

In recent years, traditional sentiment analysis technology has been widely used (Socher et al., 2013; Hamilton et al., 2016), which mainly focuses on sentiment detection. Although the current state-of-the-art sentiment analysis system can detect the polarity of text (Zhang et al., 2018) or consider fine-grained categories (a.k.a. aspects) to make predictions (Pontiki et al., 2016), the analysis of predictions and interpretations of its causes are still limited. Lately, a large amount of work introduces human motivations into sentiment analysis and action analysis (Rashkin et al., 2018a,b; Sap et al., 2019a,b; Peng et al., 2022a). However, the aforementioned works focus on the analysis of the relationship between "motivations and actions" or "emotions and actions", without modeling a unified consideration of the relationship among motivations, emotions and actions.

[†]Corresponding author.
[‡]We will make our dataset and code publicly available at https://github.com/IndexFziQ/COMMA.

Proceedings of the 29th International Conference on Computational Linguistics, pages 163–177
October 12–17, 2022.

Researches on human activities have increased over the past two decades with many fields contributing including psychology, computer science and so on. In this paper, we focus on works in two human mental states, motivations and emotions, that drive human activities. As for motivation, psychologist Hull (Hull, 1974) believes that motivation is the drive for human actions and explains why people initiate, continue or terminate a certain action at a particular time. From area of emotion, numerous theories (Cacioppo and Gardner, 1999; Kagan, 2007; Smith, 2016) that attempt to explain the origin, function, and other aspects of emotions have fostered more intense research on emotion topic. Psychologist Plutchik (Plutchik, 1980) establishes a general psycho-evolutionary theory of emotion, which introduces eight specific distinct basic emotions. Each of basic emotions represents adaptation to a prototypical task in human activity.

Aiming at modeling the relationship among human motivations, emotions and actions in language-based human individual activities, we propose a general **Co**gnitive Fra**m**ework of Hu**m**an **A**ctivities (**COMMA**). These relationships will help the researchers of NLP areas track the cause of people's emotions and actions, and give a more reasonable explanation and analysis for results. To verify the effectiveness of our framework, we propose three NLP understanding/generation tasks, including emotion understanding, motivation understanding, and conditioned action generation. More concretely, we construct a new dataset **HAIL** (**H**uman **A**ctivities **I**n **L**ife) by automatically extracting samples with complete mental state annotation from Story Commonsense (Rashkin et al., 2018a). Experimental results on NLP applications prove the effectiveness of modeling the relationship. Furthermore, our models inspired by **COMMA** can better reveal the essential relationship among motivations, emotions and actions than existing methods.

2 COMMA

Human activities are interactions between internal mind of people and the external environment. In this part, we will describe the basic elements of **COMMA** and modeling the relationship among elements in details.

2.1 Basic Elements in COMMA

Motivations are the innate physical or psychological drives of human beings, and are the origin of human activities. Different psychological theories have different classification rules for human motivations. We utilize *hierarchy of needs* of Maslow (1943) (***physiological needs, stability, love and belonging, esteem, self-actualization***).

Emotions are the psychological responses of motivations to the degree of satisfaction with the external environment. We employ the *wheel of emotions* of (Plutchik, 1980) and use eight basic emotional dimensions (***joy, trust, sadness, surprise, fear, disgust, anger, and anticipation***). It has become a common choice in the existing emotion categorization literature (Mohammad and Turney, 2013; Zhou et al., 2016; Rashkin et al., 2018a).

Actions are people's behaviors that interact with the external environment. Actions are under the psychological condition of "one has a certain need and develops on the direction of future emotion" in this paper. Limited to the annotations of our based data Story Commonsense (Rashkin et al., 2018a), we treat story events as actions, which is a sentence in language-based form. That is, ***a story event equals to an action*** in this paper.

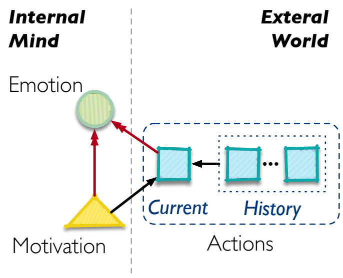
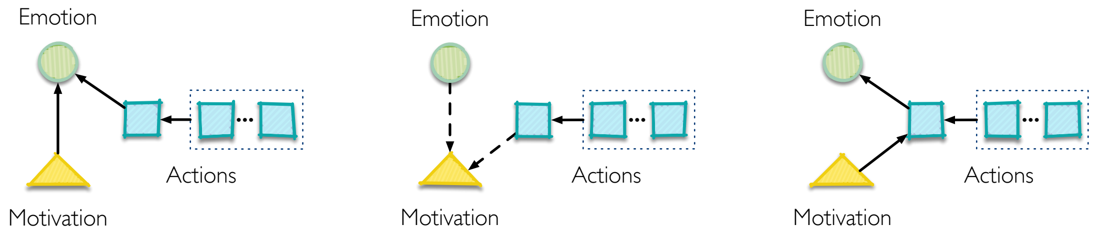


Figure 2: Relationship Modeling in **COMMA**.

2.2 Relationship Modeling

As demonstrate in Fig. 2, **COMMA** is composed of internal mind and external environment, where people own internal mind (motivations, emotions) and take actions in external environment. The solid line and the dotted line represent forward and reverse reasoning respectively. We will attempt to model relationships among the basic elements by answering the next two questions.

Q1: Where are actions from? Following the view of Hull (Hull, 1974), motivation is the drive for human actions. Intuitively, current action also caused by history actions. As shown in Fig. 2, motivation and history actions leads to the development of current action together. Meanwhile, actions develop in the direction of the future emotion. For instance, one wanted to eat, and one could eat some



(a) *Emotion Understanding* (b) *Motivation Understanding* (c) *Conditioned Action Generation*

Figure 3: Three tasks for modeling motivations, emotions, and actions in language-based human activities. The solid line and the dotted line represent forward and reverse reasoning respectively.

food and felt happy then. Conversely, one could felt sad because one had nothing to eat.

Q2: Where are emotions from? Emotions are mental states brought on by neurophysiological changes, variously associated with thoughts, feelings, behavioral responses (Panksepp, 1998; Cabanac, 2002). In simplicity, emotions come from the interaction between actions in the external world and complex mental states. In this work, we predigest this complicated process. As demonstrated in Fig. 2, emotion is conditioned by whether the action satisfy the primary motivation, similar to the statement in Li and Hovy (2017). For example, one wanted to eat, and one would happy if he ate some food, either sad if no restaurant opened.

All in all, relationships of motivations, emotions and actions are demonstrated in Fig. 2:
(1) Motivation and history actions cause action;
(2) Emotion is effect of motivation and action.

These relationships will help the researchers track human's motivations, emotions and actions.

3 Tasks and Data

To verify the effectiveness of COMMA, we propose emotion understanding, motivation understanding, and conditioned action generation tasks. Correspondingly, we build a HAIL dataset by automatically extracting samples with complete mental state annotation from Story Commonsense (Rashkin et al., 2018a).

3.1 Task Definition

The annotations of all elements in COMMA is defined as follows:
$\mathcal{A}$: The current action.
$\mathcal{H}$: The history actions.
$\mathcal{C}$: The character of current action.
$\mathcal{M}$: The motivation to drive the current action.
$\mathcal{E}$: The resulted emotion of the current action.

Emotion Understanding (EU) We formulate emotion understanding as sequence classification problems consisting of history actions, current action and people's motivations as context and a objective resulted emotion. As shown in Fig. 3(a), given the motivation $\mathcal{M}$, character $\mathcal{C}$ and all actions ($\mathcal{H}$ and $\mathcal{A}$), the EU task is to select the most plausible emotion $\mathcal{E}$.

Motivation Understanding (MU) As demonstrated in Fig. 3(b), compared with EU, motivation understanding is a reversed reasoning process. Given the emotion $\mathcal{E}$, character $\mathcal{C}$ and all actions ($\mathcal{H}$ and $\mathcal{A}$), the MU task is to reversely reason about the most plausible motivation $\mathcal{M}$.

Conditioned Action Generation (CAG) From Fig. 3(c)), CAG is the task of generating a valid action $\mathcal{A}$ and predicting the desired emotion $\mathcal{E}$ conditioned on the history actions $\mathcal{H}$, character $\mathcal{C}$ and motivation $\mathcal{M}$. Formally, the task requires to maximize $P(\mathcal{A}, \mathcal{E}|\mathcal{H}, \mathcal{C}, \mathcal{M})$.

3.2 Data Collection

To verify the effectiveness of COMMA, we construct a new dataset HAIL (**H**uman **A**ctivities **I**n **L**ife) for the above three tasks by automatically extracting from the existing resource, Story Commonsense(Rashkin et al., 2018a). Story Commonsense dataset manually annotates human motivations and emotions of the event in daily commonsense stories. It is an important resource for studying the causality of motivations, actions, and emotions in language-based individual activities. Note that, the characters of the actions (story events in this paper) are required to have both motivation and emotion labels in our collected data HAIL. In order to obtain such *(motivation, action, emotion)* samples, we align the motivation prediction and emotion prediction data sets of Story Commonsense guided by the story id and the character of the current story

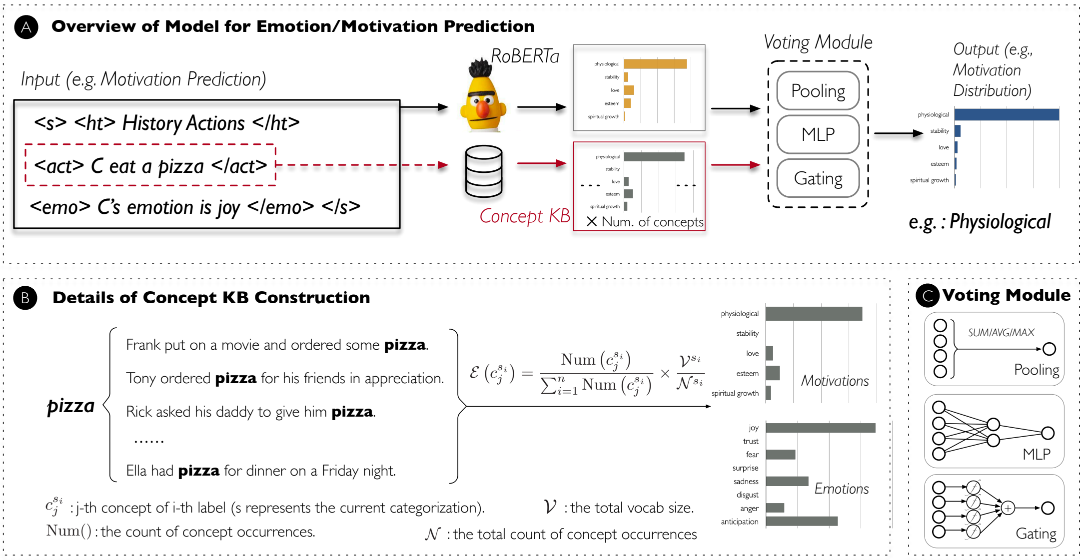


Figure 4: Model overview (A) for emotion and motivation understanding tasks (emotion understanding as an example). B is the detail of concept knowledge base construction. C shows three options of voting module in A.

event. In all, we extract 13,568 examples from Story Commonsense that meet our requirements. The I/O of three tasks are summarized in Table 1. More details, please refer to Appendix A.2.

EU Task		MU Task		CAG Task	
Input	output	Input	output	Input	output
$\mathcal{H}$,$\mathcal{A}$,$\mathcal{C}$,$\mathcal{M}$	$\mathcal{E}$	$\mathcal{H}$,$\mathcal{A}$,$\mathcal{C}$,$\mathcal{E}$	$\mathcal{M}$	$\mathcal{H}$,$\mathcal{C}$,$\mathcal{M}$	$\mathcal{A}$,$\mathcal{E}$

Table 1: Input and Output of tasks in **HAIL**.

4 Methodology

4.1 Model for Emotion Understanding

Our method combines Human Activity Encoder, Concept Knowledge Base with a voting module component, which is shown in Fig. 4. All inputs are refactored by prompt templates and special tokens (Appendix A.1) to improve understanding.

4.1.1 Human Activity Encoder

Here, we use ROBERTA (Liu et al., 2019) as our language-based human activity encoder. It is a improved robust BERT (Devlin et al., 2019) which shows state-of-the-art results in many NLP tasks. We use the hidden state representation of <s> as the sentence representation h_s.

4.1.2 Concept Knowledge Base

For emotion understanding and motivation understanding tasks, we introduce knowledge bases to calculate the distribution of commonsense knowledge in language-based actions of all motivation/emotion categories. In this paper, commonsense knowledge means commonsense concepts (i.e., words) with significant meanings that appear in language-based actions.

We build the Motivation Concept Knowledge Base (**MCKB**) in three steps. Firstly, we extract representative commonsense concepts (details in Appendix A.3). Then, we count the number of occurrences of each commonsense concepts in the category of motivations. The last step is to calculate word frequency. The knowledge distribution of each concept is computed as below:

$$\mathcal{E}\left(c_j^{s_i}\right) = \frac{\mathrm{Num}\left(c_j^{s_i}\right)}{\sum_{i=1}^{n}\mathrm{Num}\left(c_j^{s_i}\right)} \times \frac{\mathcal{V}^{s_i}}{\mathcal{N}^{s_i}} \quad (1)$$

where $c_j^{s_i}$ is j-th concept of i-th label (s represents the current categorization), Num is the number of concept $c_j^{s_i}$ occurrences. $\mathcal{V}$ is the size of the concepts vocabulary. $\mathcal{N}$ is the total number of all concept occurrences.

4.1.3 Classifier with Voting Gate

For emotion understanding task, we respectively calculate neural distribution of ROBERTA and knowledge distribution of MCKB. Lastly, we utilize a voting gate module to vote and integrate these two distributions.

Neural Distribution of Encoder Once the sentence encoding h_s is extracted, we then compute a probability distribution over labels, P_z, by the hidden representation from the classifier token

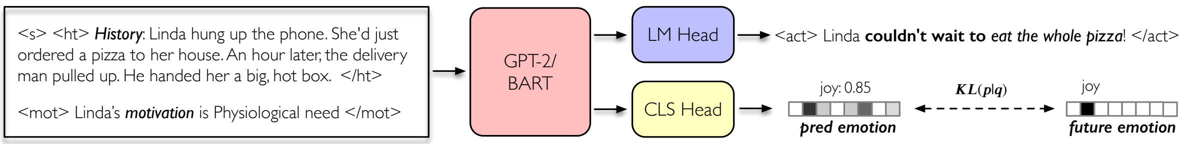


Figure 5: An example and the model architecture (GPT-2 or BART with language model head and emotion prediction head) for conditioned action generation task. Human mental states and key words in action are *italic*. The words expressing future emotion in action are **bolden**.

$h_s \in \mathbb{R}^H$ through an MLP:

$$P_z = W_2 \tanh\left(W_1 h_s + b_1\right) \tag{2}$$

where $W_1 \in \mathbb{R}^{H\times H}$, $b_1 \in \mathbb{R}^H$ and $W_2 \in \mathbb{R}^{N\times H}$, N is the number of labels. The model's predicted answer corresponds to the label of motivations with the highest probability.

Knowledge Distribution of KBs First, we use NLP parsing methods to extract representative commonsense concepts corresponding to the current action. Second, we use each commonsense concept to retrieve the corresponding distribution in concept KBs. In this way, the distribution of all commonsense knowledge $\{P_{c_1}, P_{c_2}, \ldots, P_{c_n}\}$ in the current motivation category is obtained. More details, please refer to Appendix A.3.

Voting Gate As shown in Fig. 4, the specific voting method is as follows:

$$P_f = \mathcal{F}_v(P_z, [P_{c_1}, P_{c_2}, ..., P_{c_n}]) \tag{3}$$

where n is the number of related concepts to action. Among them, $\mathcal{F}_v$ denotes voting ensemble by pooling (such as AVER, MAX, and SUM pooling), multi-layer perceptron (MLP) or gating mechanisms. Finally, the selected label of the largest probability is used as the final prediction result.

4.2 Model for Motivation Understanding

Similar to the model for emotion understanding, we build emotion concept knowledge base (**ECKB**). The difference is the categorization and the dimension of **ECKB**. The remaining modules are the same as Emotion Understanding.

4.3 Model for Conditioned Action Generation

As shown in Fig. 5, for conditioned action generation task, we employ pre-trained transformer (Vaswani et al., 2017) based language models (LM) because of their exceptional performance across related NLG tasks (Forbes et al., 2020; Rudinger et al., 2020; Sakaguchi et al., 2020; Peng et al., 2022b). Specifically, we select two standard text generation models for conditioned action generation task, BART and GPT2. **1. BART** (Lewis et al., 2020) is an encoder-decoder architecture; **2. GPT2** (Radford et al., 2019) is a single "standard" LM. We call the models trained in our settings as **COG-BART** and **COG-GPT2** for performing experiments. Besides, in order to control the direction of action generation is oriented to the future emotion, we adopt a emotion predictor to minimum the distance between given emotion of action and the given emotion label. To teach the model semantic information of the input text, which are motivations and all actions, we design prompt template for action generator (Table 9 in Appendix A.1).

4.4 Training

Emotion Understanding and Motivation Understanding For encoders of these two tasks, we adopt the general MLP classification head and obtain the distribution on each label after fine-tuning with cross-entropy loss:

$$\mathcal{L}_{\text{CLS}} = -\sum_{i=1}^{n} p\left(x_i\right) \log\left(q\left(x_i\right)\right) \tag{4}$$

where x_i represent one sample in EU and MU tasks.

Conditioned Action Generation GPT-2 and BART is trained to learn to produce the action $\mathcal{A}$ of the given history actions $\mathcal{H}$, motivation $\mathcal{M}$ and corresponding character C. To achieve this goal, our approach is trained to maximize the conditional log-likelihood of predicting the object tokens of $\mathcal{A}$:

$$\mathcal{L}_{\text{LM}} = -\sum_{t=|\mathcal{H}|+|\mathcal{M}|+|\mathcal{C}|}^{|\mathcal{H}|+|\mathcal{M}|+|\mathcal{C}|+|\mathcal{A}|} \log P\left(x_t \mid x_{<t}\right) \tag{5}$$

What's more, the predicted emotion distribution of emotion predictor is supervised by the given emotion label distribution with KL-divergence.

$$\mathcal{L}_{\text{KL}} = \mathbf{KL}(p(e_i)|q(e_i)) \tag{6}$$

where $p(e_i)$ is the predicted emotion distribution, and $q(e_i)$ is the given emotion label distribution.

To summarize, the total loss is:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{LM}} + \lambda_2 \mathcal{L}_{\text{KL}} \tag{7}$$

where λ_* is the hyper-parameter controlling the proportion of each part.

5 Experimental Setup

5.1 Baselines

We select the following three baselines for emotion understanding and motivation understanding tasks: **1. GRU** (Chung et al., 2014) is a one-layer bi-GRU encodes the input text and concatenates the final time step hidden states from both directions to yield the sentence representation h_s. **2. BERT** (Devlin et al., 2019) is a standard pre-trained language model. We concatenate sentences using specific separator tokens (`[CLS]` and `[SEP]`). Finally, we take the hidden state representation of `[CLS]` in the last layer of BERT as the overall representation h_s of sentence pairs. **3. RoBERTa** (Liu et al., 2019) is a improved robust BERT which shows state-of-the-art results in many NLP tasks. We use the hidden state representation of <s> as the sentence representation h_s.

For conditioned action generation task, we choose GPT2 (Radford et al., 2019) fine-tuned on ROCStories (Mostafazadeh et al., 2016) and **HAIL** as baselines.

5.2 Implement Details

We train baselines and our models on 9k **HAIL** training examples, then select hyper-parameters based on the best performing model on the dev set (2k), and then report results on the test set (2k). We employ GPT2 large (1.5B), BART large (680M) and ROBERTA large (340M) for our model. We implement our methods with HuggingFace[§] (Wolf et al., 2020) PyTorch (Paszke et al., 2019). We use V-100 GPU to run the experiments. More details, refer to Appendix A.4.

5.3 Metrics

Automatic Metrics. We report the micro-averaged precision (**P**), recall (**R**), and **F1** score[¶] for emotion and motivation understanding tasks.

For conditioned action generation task, we adopt three automatic measures to evaluate the generated textual action distribution both on content quality and rationality. We use the following measures: **(1)** Perplexity (**PPL**) as an indicator of fluency. A smaller value is better. **(2) BLEU** (Papineni et al., 2002) score with n is 1, 2, 4. **(3) Rouge** (Li et al., 2016) score with n is 1, 2 or L.

Human Evaluation Metrics. We also conduct a human evaluation of generated action. Crowd-workers are required to evaluate actions on a 0-3 scale (3 being very good) from two different perspectives: (1) content quality to indicate whether the generated action is **fluent and coherent**, and (2) content rationality to assess whether it **follows the given motivations and emotions**.

5.4 Results of Automatic Evaluation

Emotion Understanding. We show results on the test set in Table 2. Our approach, which using prompt template, constructed KBs, ROBERTA and voting module, achieves the highest score of all models. It is interesting that the emotion understanding task is hard for pre-trained language models with only 59.12 F1 score. This task would need more knowledge and reasoning abilities.

Motivation Understanding. As shown in Table 3, we can conclude that our method outperform other models. However, the improvement of our baseline in this task is small. It is possible that motivation understanding task needs the ability of reasoning. All in all, motivation understanding task is challengeable for the state-of-the-art models in natural language understanding tasks.

Conditioned Action Generation. Table 4 shows that our COG-GPT2 and COG-BART outperforms all baselines, indicating that it can serve as good base action generation model. We can conclude that the BLEU-1 score of COG-GPT2 models is the best. For Rouge score, COG-BART model shows best performance. One reason could be that the summarization task is helpful for generating text with larger recall.

Summary. In conclusion, our approach shows better performance than other implemented state-of-the-art models with the relationship of motivations, emotions and actions. The results of all tasks verify the feasibility of **COMMA**.

5.5 Results of Human Evaluation

We also performed manual evaluation for conditioned action prediction. We randomly selected 100 instances from the test set and used the evaluated model to generate actions. In our work, we compare the generated stories in pairs, and each

[§]https://github.com/huggingface/transformers
[¶]https://github.com/scikit-learn/scikit-learn

Models	P	R	F1
GRU (Chung et al., 2014)	36.23	36.76	36.51
$\text{BERT}_{\text{BASE}}$† (Devlin et al., 2019)	47.63	54.34	49.77
$\text{BERT}_{\text{LARGE}}$† (Devlin et al., 2019)	53.95	55.23	53.23
$\text{ROBERTA}_{\text{BASE}}$† (Liu et al., 2019)	51.47	55.64	53.09
$\text{ROBERTA}_{\text{LARGE}}$† (Liu et al., 2019)	54.36	58.27	55.93
OURS	**56.75**	**60.39**	**59.12**
(1) w/o ROBERTA	54.04	58.77	58.77
(2) w/o Knowledge Base	54.73	58.90	57.62
(3) w/o prompt template	54.95	58.23	56.23
(4) w/o voting module	55.81	59.01	57.54
(5) w/o motivation	38.38	52.88	45.74

Table 2: Results of Emotion Understanding. Our Approach is our proposed model. † means following the experimental settings in papers.

Models	P	R	F1
GRU (Chung et al., 2014)	40.53	40.27	40.89
$\text{BERT}_{\text{BASE}}$† (Devlin et al., 2019)	60.57	60.80	60.28
$\text{BERT}_{\text{LARGE}}$† (Devlin et al., 2019)	61.17	61.45	60.96
$\text{ROBERTA}_{\text{BASE}}$† (Liu et al., 2019)	61.53	61.62	61.06
$\text{ROBERTA}_{\text{LARGE}}$† (Liu et al., 2019)	63.50	63.97	63.57
OURS	**64.57**	**64.56**	**63.96**
(1) w/o ROBERTA	59.68	59.24	58.77
(2) w/o Knowledge Base	62.34	62.35	62.57
(3) w/o prompt template	63.22	63.76	62.19
(4) w/o voting module	63.65	63.93	62.44
(4) w/o emotion	64.02	63.95	63.42

Table 3: Results of Motivation Understanding. Our Approach is our proposed model. † means following the experimental settings in papers.

pair is evaluated by 3 judges. The last two columns of Table 4 report the average improvements as well as absolute scores for content quality and rationality. We can conclude that models with our designed prompt template and training loss outperform the models pre-trained on story corpus with the language model objective. It is interesting that the content of generated actions is fluent and grammatical, which indicates that GPT2 and BART is good at organize natural language.

6 Analysis and Discussion

6.1 Ablation Study

To analyze the importance of different modules in our baseline models, we perform ablation study on our approach in emotion and motivation understanding. As shown in Table 2 and Table 3, (1) denotes that the semantic representation of ROBERTA is crucial for understanding tasks. Compared (1) and (2), we find that ROBERTA and KBs have the similar scores in motivation understanding. (3) and (4) indicate the importance of prompt template and voting modules designed for ROBERTA. (5) strength that joint modeling motivation, emotion and action is helpful for emotion and motivation understanding.

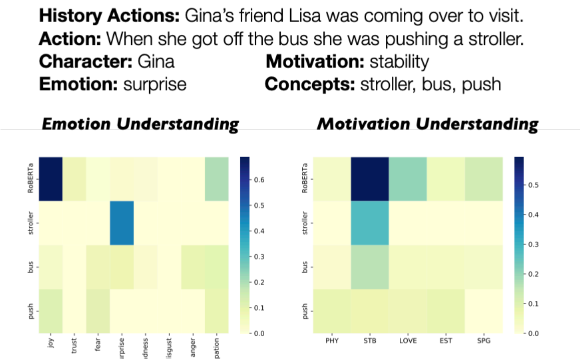


Figure 6: Case study for emotion and motivation understanding. Our framework can give better interpretability.

6.2 Human A/B Test

Human A/B test is also conducted. We try to directly compare our model with other baselines. We randomly sample 100 examples each for our model and baseline models. Three annotators are given generated responses from either our model or baselines in random order and are asked to choose a better one. They can either choose one of the responses or select "Tie" when the quality of provided options are hard to access. Results in Table 5 confirm that the responses from **COMMA** are more preferred by human judges.

6.3 Case Study

Emotion and Motivation Understanding. Fig. 6 illustrates the distribution of ROBERTA and our concept knowledge base. Our method can bring with better interpretability with the knowledge of key words in human activities. Fig. 6 shows that knowledge base predicts correctly in emotion understanding and help motivation understanding.

Conditioned Action Generation. Since the proposed models can generate actions conditioned on one's motivation, they can be used to unfold action in diverse situations for a combination of history actions, character, motivation, and emotion. We demonstrate this capability in Table 6. $\surd$ means reasonable. 🤦 means that the generated action can not express the corresponding aspect. 🤔 represents that the consistency is debatable. It can be concluded that **motivation and emotion are all**

Models	Automatic Eval							Human Eval	
	PPL	BLEU-1	BLEU-2	BLEU-4	Rouge-1	Rouge-2	Rouge-L	Content	Plausible
GPT2+ROC‡	12.47	17.24	6.26	2.16	6.86	0.23	6.44	2.36	0.79
GPT2+HAIL‡	11.83	16.46	5.81	1.92	7.32	0.38	6.71	2.24	0.56
COG-GPT2	6.85	22.48	**7.54**	**2.85**	10.86	1.05	10.25	2.79	**2.12**
w/o $\mathcal{E}$	7.99	22.29	7.58	2.81	10.22	0.94	9.66	2.85	1.79
w/o $\mathcal{M}$	8.56	21.71	7.12	2.48	10.57	0.96	10.06	2.72	1.63
COG-BART	**6.58**	**24.51**	2.26	0.31	**18.71**	**3.11**	**17.24**	**2.87**	1.98
w/o $\mathcal{E}$	7.65	23.62	2.01	0.22	17.68	2.74	16.25	2.86	1.58
w/o $\mathcal{M}$	8.86	23.98	1.94	0.16	18.53	2.72	16.99	2.79	1.85

Table 4: Automatic and human evaluation results of our COG-GPT2 and COG-BART models on conditioned action generation. COG-GPT2 and COG-BART are trained with the combination of character $\mathcal{C}$, motivation $\mathcal{M}$ and are required to predict emotion $\mathcal{E}$. ‡ represents pre-training GPT2 with language model objective on the corresponding corpus. ROC is ROCStories (Mostafazadeh et al., 2016) and **HAIL** is the train set of our proposed dataset.

Methods	Win	Loss	Tie	κ
COG-GPT2 v.s. GPT2+ROC	**54.2%**	19.5%	26.3%	30.8
COG-GPT2 v.s. GPT2+HAIL	**49.4%**	18.7%	31.9%	29.6
COG-BART v.s. GPT2+ROC	**53.3%**	18.4%	28.3%	28.9
COG-BART v.s. GPT2+HAIL	**54.3%**	14.6%	31.1%	31.3

Table 5: Human A/B Test of **COMMA**. Results show that **COMMA** performs baseline models sufficiently. κ denotes Fleiss' kappa (all are fair agreement or moderate agreement). The p-value of scores < 0.05 in sign test.

Models	**Given** $\mathcal{H}$ = *Kim and her glass went on a field trip to an aquarium.* $\mathcal{C}$=Kim, $\mathcal{M}$=spirit growth, $\mathcal{E}$=joy	R
GPT2+ROC	*Kim has a very good sense of humor .*	🙅‍♂️
GPT2+HAIL	*The personality of Kim could be humorous.*	🙅‍♂️
COG-GPT2	*She was able to get a job at a local restaurant.*	√
w/o $\mathcal{E}$	*She was so excited to go.*	🤔
w/o $\mathcal{M}$	*Kim was a very hard working woman.*	🙅‍♂️
COG-BART	*Kim had always wanted to go to the beach.*	√
w/o $\mathcal{E}$	*Kim had always wanted to be a pilot.*	√
w/o $\mathcal{M}$	*Kim and her friends decided to go on a date.*	🤔
Human	*Kim enjoyed looking at the sea creatures.*	√

Table 6: Case study of conditioned action generation task for all models, which are tested with history actions, character Kim, motivation spirit growth, emotion joy. Rationality is abbreviated as **R**. The colored text means generated action satisfy the aspects of mental states.

important for action generation. COG-GPT-2 tends to generate short but reasonable actions. But actions generated by COG-BART usually are long but repetitive. From the samples tagged by thinking face, we can see that only motivations or emotions are hard to make action prediction.

Furthermore, we perform some case studies with different inputs. The actions, with the inputs of specific history actions $\mathcal{H}$, character $\mathcal{C}$, motivation $\mathcal{M}$ or emotion $\mathcal{E}$, are generated by COG-GPT-2, COG-BART or human writing. From the first line of Fig. 7, *Jose has a spirit growth need* and *Jose's emotional expectation is joy*. As demonstrated by the three actions, we can conclude that the COG-GPT-2, COG-BART based models can generate reasonable actions. Interestingly, COG-BART can guess that the destination of the trip is Las Vegas, which is competitive to human writing. In the second example, all actions can not clearly express that the emotional expectation is *joy*, where *a big bowl* generated by COG-GPT-2 could somewhat show the happiness of Tom. Last but not least, *Tim was afraid to go outside* and *Tim went to the store to buy a new pair of shoes* are plausible corresponding to the stability need. It is possible that stability need is more abstract for pre-trained language model (PLM) to understand. Besides, we can find that all actions lack the expression of *trust* emotion. One reason is that the PLM based models are insensitive to emotional inputs, which is challengeable in future work.

6.4 Visualization Analysis

To verify the claim **COMMA** *can reveal the essential relationship among motivations, emotions, and actions*, we conduct a visualization analysis of relationships among motivations, actions, and emotions with our approach. Fig. 8 demonstrates the matrix of final prediction probability of motivations and emotions in emotion understanding tasks. The matrix makes motivations, actions and emotions close together and shows that motivations (spiritual growth) have the future emotion (i.e. anticipation($195.8E^{-3}$)). With this matrix, we can better reveal the essential relationship among motivations, emotions, and actions. Therefore, we can

Input	Actions	*C*	*M*	*E*
Character: Jose	*Cog-GPT-2: Jose was very excited to go on a trip.*	√	√	√
Motivation: spirit growth	*Cog-BART: Jose was excited to go on vacation to Las Vegas.*	√	√	√
Emotion: joy	*Human: He book a trip to the Yukon to fulfill his dream.*	√	√	√
Character: Tom	*Cog-GPT-2: Tom was eating a big bowl of cereal.*	√	√	🤔
Motivation: physiology	*Cog-BART: Tom went to the grocery store to buy some food.*	√	√	🙅
Emotion: joy	*Human: Marc offers to buy Tom skyline chilli if he pays him back.*	√	√	🙅
Character: Tim	*Cog-GPT-2: Tim was afraid to go outside.*	√	🤔	🙅
Motivation: stability	*Cog-BART: Tim went to the store to buy a new pair of shoes.*	√	🙅	🙅
Emotion: trust	*Human: He freaked out and cried uncontrollably.*	√	√	🤔

Figure 7: Case study with different inputs on conditioned action generation task. √ means reasonable. 🙅 means that the generated action can not express the corresponding aspect. 🤔 represents that the consistency is debatable. The colored text means generated action satisfy the aspects of mental states.

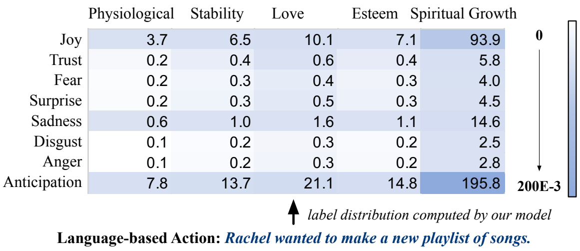


Figure 8: Visualization of final distributions of model. Each element of the above matrix uses scientific notation, and the exponent is -3.

supply more deep explanations about the relationship of motivations, actions, and emotions based on visualization analysis.

7 Related Work

There have been many large-scale language-based resources to explore human mental state, such as motivations and emotions. (Rashkin et al., 2018b; Sap et al., 2019a; Hwang et al., 2020) explore the human mental states in narrative text with series of "if-then" relationships. SOCIAL IQA was introduced by (Sap et al., 2019b) for probing emotional and social intelligence in a variety of everyday situations. Most similar to our work, Rashkin et al. (2018a) put forward Story Commonsense, which is the causal reason for the changes in the psychological state of the characters in the story.

Recently, a lot of work has begun to consider introducing various mental state of human beings into sentiment analysis and other NLP downstream tasks. Li and Hovy (2017) explore the importance of human motivations for sentiment analysis and consider emotion as a specific event or entity that realizes the mental state of human satisfaction with oneself. Otani and Hovy (2019) regard human motivation as the driving force of human emotions, and take motivation detection as the first step of emotion detection, which improves the sentiment analysis of evaluation. (Du et al., 2019; Ammanabrolu et al., 2021; Xu et al., 2020; Brahman and Chaturvedi, 2020) use the knowledge generated by COMET regarded the psychological state of COMET as a condition for story generation.

All in all, the existing language-based resources and works focus on the binary relationship between action and each mental state. Diversely, we first propose a cognitive framework that aims to analyze comprehensive relationships among motivations, emotions and actions in language-based human individual activities.

8 Conclusion and Future work

In this paper, we propose a **C**ognitive Fra**m**ework of Hu**m**an **A**ctivities (**COMMA**). To verify the effectiveness of our cognitive framework, we introduce three challenging NLP tasks, automatically construct a dataset **HAIL**, and propose the corresponding methods. Experimental results show a better understanding of the relationship among motivations, emotions and actions under our **COMMA** than existing methods. Modeling the relationship among motivations, emotions and actions in human activities can allow researchers to reason the essential causes of human activities from the cognitive perspective and supply reasonable explanations.

In future work, we will explore **COMMA** on various NLP downstream applications, such as intelligent dialogue, controllable text generation and public opinion analysis. Besides, it is interesting to further analyse more complex human activities.

Acknowledgement

We thank all anonymous reviewers for their constructive comments and useful advice. Also thanks for the discussion with Yunpeng Li, Yajing Sun, Yongxiu Xu, Ping Guo, Xinyu Zhang, Yao Dong and Yige Chen. This work is supported by the National Natural Science Foundation of China (No.62006222 and No.U21B2009). Thanks for COLING organizers and the proposed pre-trained language models, data, codes.

Contribution List Yuqiang Xie: Idea, Paper Writing, Coding; Yue Hu: Guiding, Discussion; Wei Peng: Discussion, Coding; Guanqun Bi: Discussion, Paper Polish; Luxi Xing: Review.

Thanks for the hard work and dedication of all team members.

References

Prithviraj Ammanabrolu, W. Cheung, William Broniec, and Mark O. Riedl. 2021. Automated storytelling via causal, commonsense plot ordering. In *AAAI*.

Faeze Brahman and Snigdha Chaturvedi. 2020. Modeling protagonist emotions for emotion-aware storytelling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 5277–5294. Association for Computational Linguistics.

Michel Cabanac. 2002. What is emotion? *Behavioural Processes*, 60:69–83.

John T. Cacioppo and Wendi L. Gardner. 1999. Emotion. *Annual Review of Psychology*, 50:191–214. PMID: 10074678.

Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Li Du, Xiao Ding, Ting Liu, and Zhongyang Li. 2019. Modeling event background for if-then commonsense reasoning using context-aware variational autoencoder. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 2682–2691. Association for Computational Linguistics.

Maxwell Forbes, Jena D. Hwang, Vered Shwartz, Maarten Sap, and Yejin Choi. 2020. Social chemistry 101: Learning to reason about social and moral norms. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 653–670. Association for Computational Linguistics.

William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 595–605. The Association for Computational Linguistics.

C. L. Hull. 1974. Essentials of behavior. In *New Haven: Published for the Institute of Human Relations by Yale University Press*.

Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2020. COMET-ATOMIC 2020: On symbolic and neural commonsense knowledge graphs. *CoRR*, abs/2010.05953.

Jerome Kagan. 2007. What is emotion?: History, measures, and meanings. In *Yale University Press*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7871–7880. Association for Computational Linguistics.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 110–119. The Association for Computational Linguistics.

Jiwei Li and Eduard Hovy. 2017. *Reflections on Sentiment/Opinion Analysis*, pages 41–59. Springer International Publishing, Cham.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Abraham Harold Maslow. 1943. A theory of human motivation. In *Psychological review*.

Saif Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Comput. Intell.*, 29(3):436–465.

Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James F. Allen. 2016. A corpus and cloze evaluation for deeper understanding of commonsense stories. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 839–849. The Association for Computational Linguistics.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. 2020. Adversarial NLI: A new benchmark for natural language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 4885–4901. Association for Computational Linguistics.

Naoki Otani and Eduard H. Hovy. 2019. Toward comprehensive understanding of a sentiment based on human motives. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4672–4677. Association for Computational Linguistics.

Jaak Panksepp. 1998. Affective neuroscience: The foundations of human and animal emotions. *Psychology*.

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 of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*, pages 311–318. ACL.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 8024–8035.

Wei Peng, Yue Hu, Yuqiang Xie, Luxi Xing, and Yajing Sun. 2022a. Cogintac: Modeling the relationships between intention, emotion and action in interactive process from cognitive perspective. *CoRR*, abs/2205.03540.

Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, Yajing Sun, and Yunpeng Li. 2022b. Control globally, understand locally: A global-to-local hierarchical graph network for emotional support conversation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 4324–4330. ijcai.org.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia V. Loukachevitch, Evgeniy V. Kotelnikov, Núria Bel, Salud María Jiménez Zafra, and Gülsen Eryigit. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*, pages 19–30. The Association for Computer Linguistics.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. In *OpenAI Blog*.

Hannah Rashkin, Antoine Bosselut, Maarten Sap, Kevin Knight, and Yejin Choi. 2018a. Modeling naive psychology of characters in simple commonsense stories. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 2289–2299. Association for Computational Linguistics.

Hannah Rashkin, Maarten Sap, Emily Allaway, Noah A. Smith, and Yejin Choi. 2018b. Event2mind: Commonsense inference on events, intents, and reactions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 463–473. Association for Computational Linguistics.

Rachel Rudinger, Vered Shwartz, Jena D. Hwang, Chandra Bhagavatula, Maxwell Forbes, Ronan Le Bras, Noah A. Smith, and Yejin Choi. 2020. Thinking like a skeptic: Defeasible inference in natural language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, EMNLP 2020, Online Event, 16-20 November 2020*, pages 4661–4675. Association for Computational Linguistics.

Keisuke Sakaguchi, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. 2020. Winogrande: An adversarial winograd schema challenge at scale. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8732–8740. AAAI Press.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, and Yejin Choi. 2019a. ATOMIC: an atlas of machine commonsense for if-then reasoning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 3027–3035. AAAI Press.

Maarten Sap, Hannah Rashkin, Derek Chen, Ronan Le Bras, and Yejin Choi. 2019b. Social iqa: Commonsense reasoning about social interactions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4462–4472. Association for Computational Linguistics.

Rishi Sharma, James Allen, Omid Bakhshandeh, and Nasrin Mostafazadeh. 2018. Tackling the story ending biases in the story cloze test. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 752–757. Association for Computational Linguistics.

Tiffany Watt Smith. 2016. The book of human emotions. In *Little, Brown, and Company*.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1631–1642. ACL.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020 - Demos, Online, November 16-20, 2020*, pages 38–45. Association for Computational Linguistics.

Peng Xu, Mostofa Patwary, Mohammad Shoeybi, Raul Puri, Pascale Fung, Anima Anandkumar, and Bryan Catanzaro. 2020. MEGATRON-CNTRL: controllable story generation with external knowledge using large-scale language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 2831–2845. Association for Computational Linguistics.

Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 8(4).

Deyu Zhou, Xuan Zhang, Yin Zhou, Quan Zhao, and Xin Geng. 2016. Emotion distribution learning from texts. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 638–647. The Association for Computational Linguistics.

A Appendix

A.1 Prompt Template for Input

In emotion understanding task, a model is given history actions $\mathcal{H}$, character $\mathcal{C}$, a label of motivations $\mathcal{M}$ and a textual action A. In motivation understanding task, inputs of model are history actions $\mathcal{H}$, character $\mathcal{C}$, a emotional label $\mathcal{E}$ and a textual action A. In conditoned action generation task, history actions $\mathcal{H}$, character $\mathcal{C}$, a label of motivations $\mathcal{M}$ and a emotional label $\mathcal{E}$ are given to the generator. We design simple prompt templates to expand the semantic information of the motivation and emotion labels, and also indicate the character owning the motivation and emotion. All templates for EU and MU tasks are as the following:

```
C's history actions are ___.
C's motivation is ___ .
C's action is ___.
C's emotion is ___.
```

Table 9 shows the template for conditioned action generation. In summary, this technique can enrich the semantic information of the labels and bring the labels with the given character's information. Ablation studies show the effectiveness of prompt template.

A.2 Data Collection

To verify the effectiveness of our cognitive framework, we construct a new dataset **HAIL** (**H**uman **A**ctivities **I**n **L**ife) for the above four tasks by automatically extracting from the existing resource, Story Commonsense(Rashkin et al., 2018a). Story Commonsense dataset manually annotates human motivations and emotions of the event in daily commonsense stories. It is an important resource for studying the causality of motivations, actions, and emotions in language-based individual activities. Note that, the actors of the actions are required to have both motivations and emotion labeling in our collected data **HAIL**. In order to obtain such *(motivation, action, emotion)* samples, we utilize NLTK[*] (a natural language processing toolkit) and design some rules. In all, we extract 13,568 examples in Story Commonsense that meet our requirements. Fig. 11 denotes the data statistics of label distributions in **HAIL**, including motivations and emotions. The label distribution is relatively uniform, which is conducive to the learning of the model.

Data Analysis We perform analysis about the gender bias of open-text actions in HAIL. As is shown Fig. 10, our dataset have a good distribution considering the gender of individual in all actions.

[*]http://www.nltk.org/

Hyper-parameter	Value
LR	{1e-5, 2e-5}
Batch size	{16, 32, 64}
Gradient norm	1.0
Warm-up	0.1
Max. input length (# subwords)	200
Epochs	{3, 5, 10}

Table 7: Hyper-parameters of models based on BERT and ROBERTA for emotion and motivation understanding tasks.

Hyper-parameter	Value
LR	1e-5
λ_1	1
λ_2	1.5
Batch size	32
Gradient norm	1.0
Warm-up	0.1
Max. input length (# subwords)	200
Max. output length (# subwords)	60
Max # Epochs	30

Table 8: Hyper-parameters of models based on BART and GPT-2 for action prediction task.

This mechanism ensures that there is a clear and agreed-upon relationship between *needs-action-emotion* in the story, and avoids subjectivity and ambiguity in SCT (Sharma et al., 2018) and certain NLU tasks (Nie et al., 2020).

A.3 Knowledge Distribution of KBs

The knowledge bases can give the knowledge distribution of the motivation/emotion category according to the commonsense concepts appeared in the action, which corresponds to *Knowledge Distribution*. Specifically, we use tools such as NLTK[†] and Spacy[‡], and then remove stop words and high-frequency words to extract representative commonsense concepts corresponding to the current action. Finally, we use each commonsense concept to retrieve the corresponding distribution in MCKB or ECKB. In this way, the distribution of all commonsense knowledge $\{P_{c_1}, P_{c_2}, \dots, P_{c_n}\}$ in the current motivation/emotion category is obtained.

Base on the training set of our proposed **HAIL**,

[†]http://www.nltk.org/

[‡]https://spacy.io/

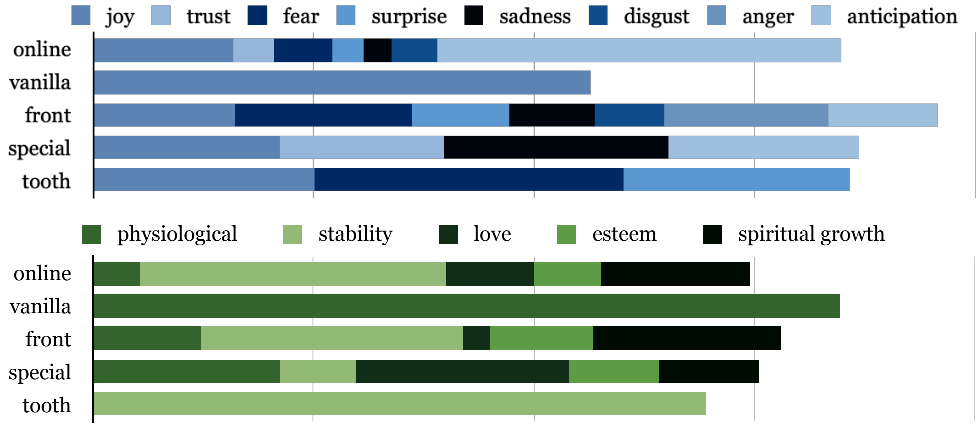


Figure 9: Examples of distribution of motivations Concept KB (below) or Emotion Concept KB (top).

Input Prompt Template	Output Prompt Template
[ht] C's history actions are __ [/ht] and [mot] C has __ motivation [/mot]	[act] __ [/act]

Table 9: In conditioned action generation task, input formats of COG-BART and COG-GPT2.

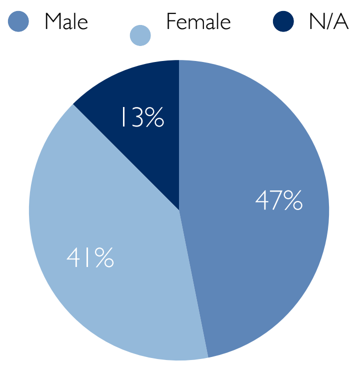


Figure 10: Analysis of gender bias of HAIL.

we automatically construct knowledge bases of motivations and emotions. Examples of them are shown individually in Fig. 9. These two KBs can be used to make prediction or assist the decision-making of the deep model, Moreover, it can also be used to evaluate or explain the forecast results.

A.4 Implement Details

We train baseline models on 9k **HAIL** training examples, then select hyper-parameters based on the best performing model on the dev set (2k), and then report results on the test set (2k). The hyper-parameters of BART and GPT-2 is shown in Table 7. The hyper-parameters of BERT and ROBERTA is shown in Table 8. We use V-100 GPU to run the experiments.

A.5 Comparison of Different Inputs

In order to analyze the effectiveness of our baseline models for action generation, we also perform some case studies with different inputs. The actions, with the inputs of specific history actions $\mathcal{H}$, character $\mathcal{C}$, motivation $\mathcal{M}$ or emotion $\mathcal{E}$, are generated by COG-GPT-2, COG-BART or human writing. From the first line of Fig. 12, *Jose has a spirit growth need* and *Jose's emotional expectation is joy*. As demonstrated by the three actions, we can conclude that the COG-GPT-2, COG-BART based models can generate reasonable actions. Interestingly, COG-BART can guess that the destination of the trip is Las Vegas, which is competitive to human writing. In the second example, all actions can not clearly express that the emotional expectation is *joy*, where *a big bowl* generated by COG-GPT-2 could somewhat show the happiness of Tom. Last but not least, *Tim was afraid to go outside* and *Tim went to the store to buy a new pair of shoes* are plausible corresponding to the stability need. It is possible that stability need is more abstract for pre-trained language model (PLM) to understand. Besides, we can find that all actions lack the expression of *trust* emotion. One reason is that the PLM based models are insensitive to emotional inputs, which is challengeable in future work.

A.6 Future work

Modeling the relationship among motivations, emotions and actions in human activities can allow re-

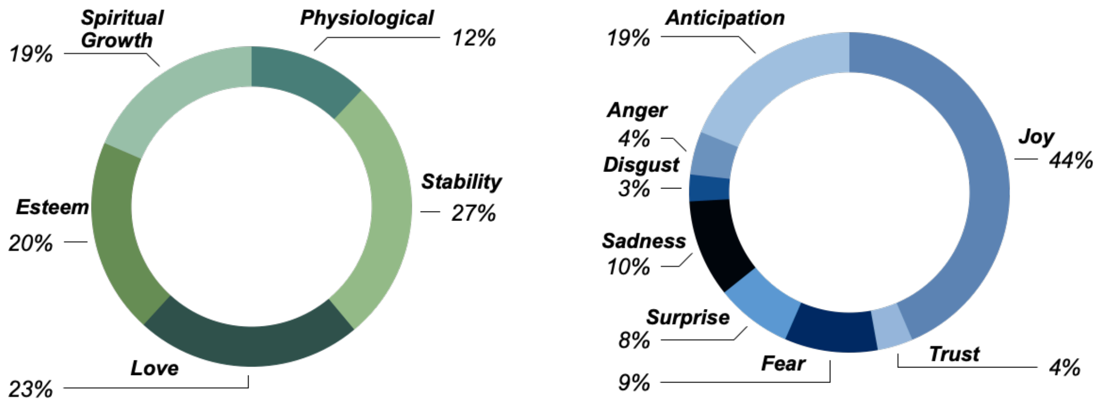


Figure 11: Data statistics of label distributions in **Hail**, including Human motivations (Left Pic) and Emotion Reactions (Right Pic).

Input	**Actions**	***C***	***M***	***E***
Character: Jose	*Cog-GPT-2: Jose was very excited to go on a trip.*	√	√	√
Motivation: spirit growth	*Cog-BART: Jose was excited to go on vacation to Las Vegas.*	√	√	√
Emotion: joy	*Human: He book a trip to the Yukon to fulfill his dream.*	√	√	√
Character: Tom	*Cog-GPT-2: Tom was eating a big bowl of cereal.*	√	√	🤔
Motivation: physiology	*Cog-BART: Tom went to the grocery store to buy some food.*	√	√	🙅
Emotion: joy	*Human: Marc offers to buy Tom skyline chilli if he pays him back.*	√	√	🙅
Character: Tim	*Cog-GPT-2: Tim was afraid to go outside.*	√	🤔	🙅
Motivation: stability	*Cog-BART: Tim went to the store to buy a new pair of shoes.*	√	🙅	🙅
Emotion: trust	*Human: He freaked out and cried uncontrollably.*	√	√	🤔

Figure 12: Case study with different inputs on conditioned action generation task. The actions are generated by Cog-Gpt-2, Cog-Bart or human writing with the specific character, motivation and emotion. √ means reasonable. 🙅 means that the generated action can not express the corresponding aspect. 🤔 represents that the consistency is debatable. The colored text means generated action satisfy the aspects of mental states.

searchers to reason the essential causes of human activities from the cognitive perspective and supply reasonable explanations. In future work, we will explore **Comma** on various NLP downstream applications, such as intelligent dialogue, controllable text generation and public opinion analysis. Besides, it is interesting to further analyse more complex human activities.