SocAoG: Incremental Graph Parsing for Social Relation Inference in Dialogues

Liang Qiu[1], Yuan Liang[2], Yizhou Zhao[1], Pan Lu[1], Baolin Peng[3],
Zhou Yu[4], Ying Nian Wu[1], Song-Chun Zhu[1]

[1]UCLA Center for Vision, Cognition, Learning, and Autonomy
[2]University of California, Los Angeles
[3]Microsoft Research, Redmond
[4]University of California, Davis
liangqiu@ucla.edu

Abstract

Inferring social relations from dialogues is vital for building emotionally intelligent robots to interpret human language better and act accordingly. We model the social network as an And-or Graph, named SocAoG, for the consistency of relations among a group and leveraging attributes as inference cues. Moreover, we formulate a sequential structure prediction task, and propose an α–β–γ strategy to incrementally parse SocAoG for the dynamic inference upon any incoming utterance: *(i)* an α process predicting attributes and relations conditioned on the semantics of dialogues, *(ii)* a β process updating the social relations based on related attributes, and *(iii)* a γ process updating individual's attributes based on interpersonal social relations. Empirical results on DialogRE and MovieGraph show that our model infers social relations more accurately than the state-of-the-art methods. Moreover, the ablation study shows the three processes complement each other, and the case study demonstrates the dynamic relational inference.[1]

1 Introduction

Social relations form the basic structure of our society, defining not only our self-images but also our relationships (Sztompka, 2002). Robots with a higher emotional quotient (EQ) have the potential to understand users' social relations better and act appropriately. Given a dialogue as context and a set of entities, the task of Dialogue Relation Extraction (DRE) predicts the relation types between the entities from a predefined relation set. Table 5 shows such an example from the dataset DialogRE (Yu et al., 2020).

Existing researches using BERT-based models (Devlin et al., 2018; Yu et al., 2020; Xue et al., 2020a) or graph-based models (Xue et al., 2020b;

---

[1]The code is released at https://github.com/Liang-Qiu/SocAoG-dialogues.

	Dialogue	
S1:	Well then we'll-we'll see you the day after tomorrow. Mom?! Dad?! What-what… what you guys doing here?!	
S2:	Well you kids talk about this place so much, we thought we'd see what all the fuss is about.	
S3:	I certainly see what the girls like coming here.	
S1:	Why?!	
S3:	The sexy blonde behind the counter.	
S1:	Gunther?!	
S2:	Your mother just added him to her list.	
S1:	What? Your-your list?	

	Argument Pair	Trigger	Relation Type
R1	(S2, S1)	dad	per:children
R2	(S3, Gunther)	sexy blonde	per:positive_impression
R3	(S3, S1)	mom	per:children
R4	(S1, S3)	mom	per:parents
R5	(S1, S2)	dad	per:parents

Table 1: A dialogue example from DialogRE (Yu et al., 2020). Trigger word annotations are not used for training, but rather for illustrating purpose only.

Chen et al., 2020) focus on identifying entities' relations from the semantics of dialogues—they utilize either the attention mechanism or a refined token graph to locate informative words (*e.g.*, "dad" and "mom") that imply the argument pairs' relations. However, there are still three missing parts in current models for social relation inference according to our observations. First, current models lack the explicit modeling of the relational consistency among a group of people—such consistency helps humans reason about the social relation of two targets by using their relations with a third person. For the example in Table 5, by knowing S2 and S3 are S1's parents and S3 is S1's mother, we can infer that S2 is S1's dad. Second, the personal attribute cues (*e.g.*, gender and profession) can also aid the relational inference but are not fully utilized. In the above example, besides inferring S3 is S1' mother according to S3's feminine attribute, we can also have a guess that Gunther is a waiter, which might be useful for the future social-relational inference. Third, since the BERT-based
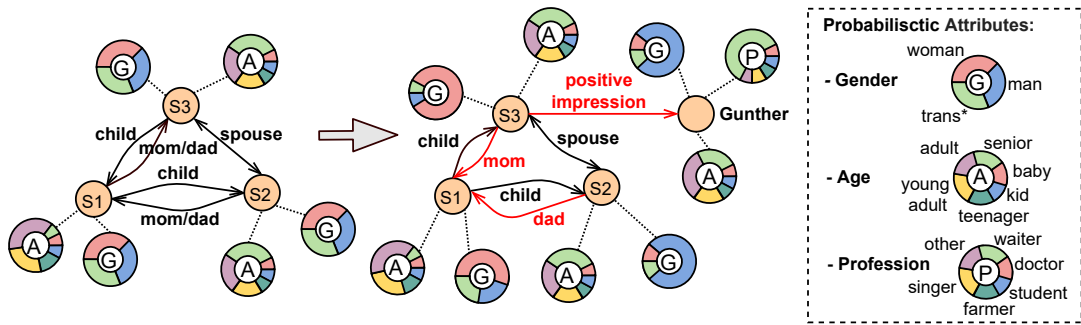
Figure 1: Our method iteratively updates the robot's belief of users' individual attributes and social relations, similar to human's reasoning process. The left and right graph show the established and updated belief, respectively.

and token-graph-based models take dialogues as a whole for relation prediction, they cannot perform dynamic inference—updating the relational belief with an incoming dialogic utterance. This can limit their ability to track the evolving relations along social interactions, *e.g.*, strangers become friends over a good chat (Kukleva et al., 2020), unveiling intermediate reasoning results, or dealing with long dialogues.

Motivated by these observations, we propose to model social relation as an attributed And-Or graph (AoG) (Zhu et al., 1998; Zhu and Mumford, 2007; Wu and Zhu, 2011; Shu et al., 2016; Qi et al., 2018), named SocAoG, and develop an incremental graph parsing algorithm to jointly infer human attributes and social relations from a dialogue. In specific, SocAoG describes social relations and personal attributes with contextual constraints of groups and hierarchical representations. To incrementally parse SocAoG and track social relations, we apply Markov Chain Monte Carlo (MCMC) to sample from the posterior probability calculated by three complementary processes (α–β–γ) (Qu et al., 2020; Zayaraz et al., 2015). Figure 1 schematically demonstrates a graph update of both relations (*i.e.*, disambiguating mom/dad and adding a new party) and attributes (*e.g.*, gender and profession) with the utterance "**S2**: *Your mother just added him to the list.*" from the example dialogue in Table 5.

We evaluate our method on two datasets of DialogRE (Yu et al., 2020) and MovieGraph (Vicol et al., 2018) for relation inference, and the results show that our method outperforms the state-of-the-art (SoTA) ones. Overall, we make the following contributions: (*i*) We propose to model and infer social relations and individual's attributes jointly with SocAoG for the consistency of attributes and social relations among a group. To the best of our knowledge, it is the first time done in the dialogue do-

main; (*ii*) The MCMC sampling from α–β–γ posterior enables dynamic inference—incrementally parsing the social relation graph, which can be useful for tracking relational evolution, reflecting the reasoning process, and handling long dialogues; (*iii*) We perform an ablation study on each process of α–β–γ to investigate the information contribution, and perform case studies to show the effectiveness of our dynamic reasoning.

2 Related Work

We review the related works on the social relation inference from documents, which is a well-studied task, and those from dialogues, which is the emerging task that our work is focused on.

2.1 Relation Inference from Documents

Most of the existing literature focus on relation extraction from professional edited news reports or websites. They typically output a set of "subject-predicate-object" triples after reading the entire document (Bach and Badaskar, 2007; Mintz et al., 2009; Kumar, 2017). While early works mostly utilize feature-based methods (Kambhatla, 2004; Miwa and Sasaki, 2014; Gormley et al., 2015) and kernel-based methods (Zelenko et al., 2003; Zhao and Grishman, 2005; Mooney and Bunescu, 2006), more recent studies use deep learning methods such as recurrent neural networks or transformers (Kumar, 2017). For example, Zhou et al. (2016) propose bidirectional LSTM model to capture the long-term dependency between entity pairs, Zhang et al. (2017) present PA-LSTM to encode global position information, and Alt et al. (2019); Papanikolaou et al. (2019) fine-tune pre-trained transformer language models for relation extraction.

Two streams of work are closely related to our method. Regarding social network modeling, while most works treat pairs of entities isolated (Yu et al.,

2020; Xue et al., 2020b; Chen et al., 2020), Srivastava et al. (2016) formulate the interpersonal relation inference as structured prediction (Belanger and McCallum, 2016; Qiu et al., 2020; Zhao et al., 2020), inferring the collective assignment of relations among all entities from a document (Li et al., 2020; Jin et al., 2020). Regarding relation evolution, a few works are aimed to learn the dynamics in social networks, *i.e.,* the development of relations, from narratives by Hidden Markov Models (Chaturvedi et al., 2017), Recurrent Neural Networks (Kim and Klinger, 2019), deep recurrent autoencoders (Iyyer et al., 2016). Our method differs from the aforementioned works by modeling the structured social relations and their changes concurrently, which can be useful for the task of tracking social network evolution (Doreian and Stokman, 1997) and unveiling the reasoning process of relations. We achieve this by parsing the graph incrementally per utterance with the proposed α–β–γ strategy.

2.2 Relation Inference from Dialogues

Recently, Yu et al. (2020) introduce the first human-annotated dialogue-based relation extraction dataset DialogRE, in which relations are annotated between arguments that appear in a dialogue session. Compared with traditional relation extraction tasks, DialogRE emphasizes the importance of tracking speaker-related information within the context across multiple sentences. SoTA methods can be categorized into token-graph models and pre-trained language models. For typical token-graph models, Chen et al. (2020) present a token graph attention network, and Xue et al. (2020b) further generate a latent multi-view graph to capture relationships among tokens, which is then refined to select important words for relation extraction. For pre-trained models, Yu et al. (2020) evaluate a BERT-based baseline model (Devlin et al., 2018) and a modified version BERTs, which takes speaker arguments into consideration. Xue et al. (2020a) propose a simple yet effective BERT-based model, SimpleRE, that takes a novel input format to capture the interrelations among all pairs of entities.

Both categories of SoTA models take a discriminative approach, whereas ignoring two key constraints on relations: *(i)* social relation consistency in a group and *(ii)* human attributes. Different from them, our method formulates the task as dialogue generation from an attributed relation graph, so

that the posterior relation estimation models both two constraints. Moreover, SoTA models also assume the relations are static—they cannot learn the dynamics of the relations, while the incremental graph updating strategy naturally enables the dynamic relation inference.

3 Problem Formulation

Our goal is to construct a social network through utterances in dialogue. The network is a heterogeneous physical system (Yongqiang et al., 1997) with particles representing entities and different types of edges representing social relations. Each entity is associated with multiple types of attributes, while each type of relation is governed by a potential function defined in human attribute and value space, acting as the social norm. The relations are often asymmetric, *e.g.*, A is B's father does not mean B is A's father. To model the network, we utilize an attributed And-Or Graph (A-AoG), a probabilistic grammar model with attributes on nodes. Such design takes advantage of the reconfigurability of its probabilistic context-free grammar to reflect the alternative attributes and relations, and the contextual relations defined on Markov Random Field to model the social norm constraints.

The social network graph, named SocAoG, is diagrammatically shown in Figure 2. Formally, SocAoG is defined as a 5-tuple:

$$\mathcal{G} =< S, V, E, X, P > \qquad (1)$$

, where S is the root node for representing the interested society. $V = V_{and} \cup V_{or} \cup V_T^e \cup V_T^a$ denotes all nodes' collection. Among them, And-nodes V_{and} represent the set of social communities, which can be decomposed to a set of entity terminal nodes, V_T^e, representing human members. Community detection is based on the social network analysis (Bedi and Sharma, 2016; Du et al., 2007), and can benefit the modeling of loosely connected social relations. Each human entity is associated with an And-node that breakdowns the attributes into subtypes such as gender, age, and profession. All the subtypes consist of an Or-node set, V_{or}, for representing branches to alternatives of attribute values. Meanwhile, all the attribute values are represented as a set of terminal nodes V_T^a. We denote E to be the edge set describing social relations, $X(v_i)$ to be the attributes associated with node v_i, and $X(\vec{e}_{ij})$ to be the social relation type of edge $\vec{e}_{ij} \in E$.
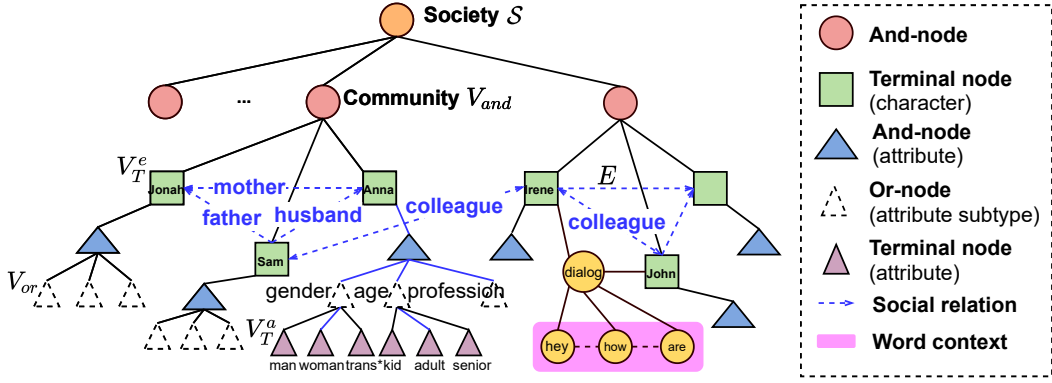
Figure 2: **SocAoG**: Attributed And-Or Graph representation of a social network. A parse graph determining each attribute and relation type is marked in blue lines. Dialogues are governed by the word context and associated human attributes and relations.

Given P to be the probability model defined on SocAoG, a parse graph pg is an instantiation of SocAoG with determined attribute selections for every Or-node and relation types for every edge. For a dialogue session with T turns $D_T = \{D^{(1)}, D^{(2)}, ..., D^{(T)}\}$, where $D^{(t)}$ is the utterance at turn t, our method infers the attributes and social relations incrementally over turns:

$$\mathcal{G}_T = \{pg^{(1)}, pg^{(2)}, ..., pg^{(T)}\} \qquad (2)$$

, where $pg^{(t)}$ represents the belief of SocAoG at the dialogue turn t. We incrementally update the pg by maximizing the posterior probability:

$$pg^* = \arg\max_{pg} p(pg|D; \theta) \qquad (3)$$

, where pg^* is the optimum social relation belief, and θ is the set of model parameters.

4 Algorithm

4.1 α–β–γ for Graph Inference

For simplicity, we denote $X(v_i)$ as $\mathbf{v}_i$ and $X(\vec{e}_{ij})$ as $\mathbf{e}_{ij}$ in the rest of the paper. We introduce three processes, *i.e.*, α, β, and γ process, to infer any SocAoG belief pg^*. We start by rewriting the posterior probability as a Gibbs distribution:

$$p(pg|D; \theta) \propto p(D|pg; \theta)p(pg; \theta)$$
$$= \frac{1}{Z} exp\{-\mathcal{E}(D|pg; \theta) - \mathcal{E}(pg; \theta)\}$$
$$(4)$$

, where Z is the partition function. $\mathcal{E}(D|pg; \theta)$ and $\mathcal{E}(pg; \theta)$ are dialogue- and social norm-based energy potentials respectively, measuring the cost of assigning a graph instantiation.

Denoting a dialogue as a sequence of words: $D = \{w_1, w_2, ..., w_{\mathcal{T}}\}$, the dialogue likelihood energy term $\mathcal{E}(D|pg; \theta)$ can be expressed with a language model conditioned on the parse graph:

$$\mathcal{E}(D|pg; \theta) = \sum_{t=1}^{\mathcal{T}} \mathcal{E}(w_t|\mathbf{c}_t, pg)$$
$$= \sum_{t=1}^{\mathcal{T}} -\log(p(w_t|\mathbf{c}_t, pg)) \qquad (5)$$

, where $\mathbf{c}_t = [w_1, ..., w_{t-1}]$ is the context vector. Intuitively, the word selection depends on the word context, the entities' attributes and their interpersonal relations. We approximate the likelihood by finetuning a BERT-based transformer with a customized input format $\langle$[CLS]D[SEP]$v_{i_0}\mathbf{e}_{i_0 j_0}v_{j_0}...v_{i_n}\mathbf{e}_{i_n j_n}v_{j_n}v_0\mathbf{v}_0...v_n$ $\mathbf{v}_n$[SEP]$\rangle$, which is a concatenation of the dialogue history D and a flattened parse graph string encoding the current belief. We call the estimation of pg from the dialogue likelihood $p(w_t|\mathbf{c}_t, pg)$ to be the α **process**. α process lacks the explicit constraints for social norms related to interpersonal relations and human attributes.

For the social norm-based potential, we design it to be composed of three potential terms:

$$\mathcal{E}(pg; \theta) = -\beta \sum_{v_i, v_j \in V(pg)} \log(p(\mathbf{e}_{ij}|\mathbf{v}_i, \mathbf{v}_j))$$
$$- \gamma_l \sum_{\vec{e}_{ij} \in E(pg)} \log(p(\mathbf{v}_i|\mathbf{e}_{ij}))$$
$$- \gamma_r \sum_{\vec{e}_{ij} \in E(pg)} \log(p(\mathbf{v}_j|\mathbf{e}_{ij}))$$
$$(6)$$

, where $V(pg)$ and $E(pg)$ are the set of terminal nodes and relations in the parse graph, respectively.

661

We call the term $p(\mathbf{e}_{ij}|\mathbf{v}_i, \mathbf{v}_j)$ the β **process**, in which we bind the attributes of node v_i and v_j to update their relation edge $\mathbf{e}_{ij}$, in order to model the constraint on relations from human attributes. Reversely, we call the terms $p(\mathbf{v}_i|\mathbf{e}_{ij})$ and $p(\mathbf{v}_j|\mathbf{e}_{ij})$ the γ **process**, in which we use the social relation edge $\mathbf{e}_{ij}$ to update the attributes of node v_i and v_j. This models the impact of relation to the attributes of related entities. β, γ_l, and γ_r are weight factors balancing α, β and γ processes. Figure 3(a) shows the graph inference schema with the three processes. Combining equation 4, 5, and 6, we get a posterior probability estimation $p(pg|D; \theta)$ of parse graph pg, with the guarantee of the attribute and social norm consistencies.
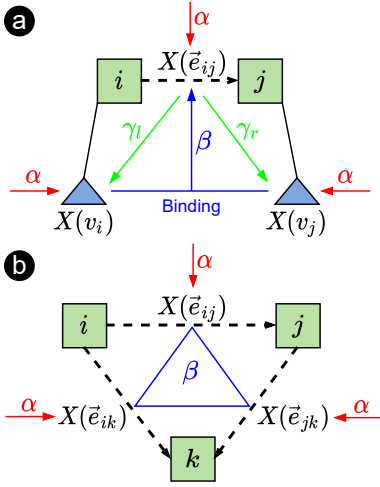

Figure 3: (a) α–β–γ process for SocAoG. (b) α–β process for reduced SocAoG without attributes. Note that this β is only modeling the interrelations among $X(\vec{e})$.

Here we also provide a reduced version of our model, SocAoG$_{reduced}$, which applies when characters' attributes annotation are not available for training[2]. With the same dialogue-based energy potential, We define the parse graph prior energy over a set of relation triangles:

$$\mathcal{E}(pg; \theta) = -\beta \sum_{\vec{e}_{ij}, \vec{e}_{ik}, \vec{e}_{jk} \in E(pg)} \log(p(\mathbf{e}_{ij}|\mathbf{e}_{ik}, \mathbf{e}_{jk})). \tag{7}$$

The method directly models the constraint of two entities' relation from their relations to others, with the inference schema demonstrated in Figure 3(b).

4.2 Incremental Graph Parsing

Incrementally parsing the SocAoG is accomplished by repeatedly sampling a new parse graph $pg^{(t)}$

---

[2]Both SocAoG and SocAoG$_{reduced}$ do not need attribute annotation during inference once trained.

Algorithm 1: Incremental SocAoG Parsing for Social Relation Inference

Input: dialogue $D_T = \{D^{(1)}, D^{(2)}, ..., D^{(T)}\}$, target argument pairs $\{a_1, a_2\}$.
Initialize $pg^{(0)}$. Initialize $\mathbf{v}_i$ and $\mathbf{e}_{ij}$.
for $t = 1, ..., T$ **do**
  for $s = 1, ..., S$ **do**
    Compute the posterior $p(pg|D^{(t)}; \theta)$.
    Make proposal moves with probabilities q_1, q_2 to get a new parse graph pg'.
    Compute the posterior $p(pg'|D^{(t)}; \theta)$.
    Compute acceptance rate $\alpha(pg'|pg, D^{(t)}; \theta)$.
    Accept/reject pg' according to the acceptance rate.
  end for
  return $\mathbf{e}_{a_1, a_2}$ from the average of accepted pg samples.
end for

---

from the posterior probability $p(pg^{(t)}|D^{(t)}; \theta)$. We utilize a Markov Chain Monte Carlo (MCMC) sampler to update our parse graph since the complexity of the problem caused by multiple energy terms.

At each dialogue turn t, we initialize the parse graph with the α classification process, by replacing all the Or-Node tokens with a special token [CLS]. We sample the parse graph for S steps and use the average value of obtained samples as an approximation of $pg^{(t)}$. We design two types of Markov chain dynamics used at random probabilities $q_i, i = 1, 2$ to make proposal moves:

- Dynamics q_1: randomly pick a relation edge $\vec{e}_{ij}$ under the uniform distribution, flip its social relation type $\mathbf{e}_{ij}$ according to the prior distribution given by β process:

$$\prod_{v_i, v_j \in V(pg)} p(\mathbf{e}_{ij}|\mathbf{v}_i, \mathbf{v}_j). \tag{8}$$

- Dynamics q_2: randomly pick a terminal node v_i and its attribute subtype under the uniform distribution, and flip the one-hot value of attribute $\mathbf{v}_i$ according to the prior distribution given by γ process:

$$\prod_{\vec{e}_{ij} \in E(pg)} p(\mathbf{v}_i|\mathbf{e}_{ij}) \prod_{\vec{e}_{ji} \in E(pg)} p(\mathbf{v}_i|\mathbf{e}_{ji}). \tag{9}$$

Using the Metropolis-Hastings algorithm (Chib and Greenberg, 1995), the proposed new parse graph pg' is accepted according to the following accep-

tance probability:

$$\alpha(pg'|pg, D; \theta) = \min(1, \frac{p(pg'|D; \theta)p(pg|pg')}{p(pg|D; \theta)p(pg'|pg)})$$
$$= \min(1, \frac{p(pg'|D; \theta)}{p(pg|D; \theta)})$$
(10)

, where the proposal probability rate is cancelled out since the proposal moves are symmetric in probability. We summarize the incremental SocAoG parsing in Algorithm 1. Dialogues give a continuously evolving energy landscape: at the beginning of iterations, $p(pg^{(0)}|D; \theta)$ is a "hot" distribution with a large energy value; by iterating the α–β–γ processes for pg updates through the dialogue, the pg converges to the pg^*, which is much cooler.

5 Experiments

5.1 Datasets

We use DialogRE (V2)[3] (Yu et al., 2020) and MovieGraph[4] (Vicol et al., 2018) for evaluating our method. Detailed descriptions on the two datasets, *e.g.*, relation and attribute types, are provided in Appendix A.

DialogRE contains 36 relation types (17 of them are interpersonal) that exist between pairs of arguments. For the joint parsing of relation and attribute, we further annotate the entity arguments with attributes from four subtypes (by following the practice of MovieGraph (Vicol et al., 2018)): gender, age, profession, and ethnicity, according to Friends Central in Fandom[5]. DialogRE is split into training (1073), validation (358), and test (357). Following previous works (Yu et al., 2020; Xue et al., 2020b), we report macro $\mathbf{F}1$ scores in both the standard and conversational settings ($\mathbf{F}1_c$).

MovieGraph provides graph-based annotations of social situations from 51 movies. Each graph comprises nodes representing the characters, their emotional and physical attributes, relationships, and interactions. We use a subset (40) of MovieGraph with available full transcripts and split the dataset into training (26), validation (6), and test (8). For MovieGraph, we only evaluate with $\mathbf{F}1$ since the trigger word annotation for computing $\mathbf{F}1_c$ is not available.

---

[3] https://github.com/nlpdata/dialogre
[4] http://moviegraphs.cs.toronto.edu/
[5] https://friends.fandom.com/wiki/Friends_Wiki

5.2 Experiment Settings

We learn the SocAoG model with a contrastive loss (Hadsell et al., 2006) comparing the posterior of a positive parse graph against a negative one. All parameters are learned by gradient descent using the Adam optimizer (Kingma and Ba, 2014). During the inference stage, for each utterance, we run the MCMC for $S = \min\{w \times (KM + K(K - 1)N), S_{max}\}$ steps given K entities, M attributes, N relations, and a sweep number of w. The probability of flipping the relation q_1 is set to 0.7 to bias towards the relation prediction at first.

5.3 Baseline Models

We compare our method with both transformer-based (**BERT**, **BERT$_S$**, **SimpleRE**) and graph-based (**GDPNet**) models. Given dialogue history D and target argument pair (v_i, v_j), **BERT** (Devlin et al., 2018) takes input sequences formatted as $\langle$[CLS]d[SEP]v_i[SEP]v_j[SEP]$\rangle$. **BERT$_S$** (Yu et al., 2020) is a speaker-aware modification of BERT, which also takes speaker information into consideration by converting it into a special token. **SimpleRE** (Xue et al., 2020a) models the relations between each pair of entities with a customized input format. **GDPNet** (Xue et al., 2020b) takes in token representations from BERT and constructs a multi-view graph with a Gaussian Graph Generator. The graph is then refined through graph convolution and DTWPool to identify indicative words.

5.4 Performance Comparison

Table 2 shows the performance comparison between different methods on the two datasets. It clearly shows that both of our models, SocAoG and SocAoG$_{reduced}$, outperform the existing methods by all the metrics. In specific, without using any additional information of attributes, SocAoG$_{reduced}$ surpasses the state-of-the-art method (SimpleRE) by 1.9% ($\mathbf{F}1$)/2.1% ($\mathbf{F}1_c$) on DialogRE testing set, and by 5.1% ($\mathbf{F}1_c$) on MovieGraph testing set. Such improvement shows the importance of relational consistency for the modeling, and proves the effectiveness of our SocAoG formulation to introduce the social norm constraints.

Moreover, by comparing between SocAoG and SocAoG$_{reduced}$, we see that SocAoG further improves most of the metrics by leveraging the attribute information for relation reasoning, *e.g.*, 69.1% *vs.* 68.6% for DialogRE testing $\mathbf{F}1$ and

| | DialogRE (V2) | | | | MovieGraph | |
| | Dev | | Test | | Dev | Test |
Methods	$\mathbf{F}1(\sigma)$	$\mathbf{F}1_c(\sigma)$	$\mathbf{F}1(\sigma)$	$\mathbf{F}1_c(\sigma)$	$\mathbf{F}1(\sigma)$	$\mathbf{F}1(\sigma)$
BERT (Devlin et al., 2018)	59.4 (0.7)	54.7 (0.8)	57.9 (1.0)	53.1 (0.7)	50.6 (1.2)	53.6 (0.3)
BERT$_S$ (Yu et al., 2020)	62.2 (1.3)	57.0 (1.0)	59.5 (2.1)	54.2 (1.4)	50.7 (1.1)	53.6 (0.4)
GDPNet (Xue et al., 2020b)	67.1 (1.0)	61.5 (0.8)	64.3 (1.1)	60.1 (0.9)	53.1 (1.1)	56.4 (0.8)
SimpleRE (Xue et al., 2020a)	68.2 (1.1)	63.4 (0.6)	66.7 (0.7)	63.3 (0.9)	55.2 (0.5)	58.1 (0.7)
SocAoG$_{reduced}$ (our method)	69.1 (0.4)	65.7 (0.5)	68.6 (0.9)	65.4 (1.1)	**60.7 (0.4)**	63.2 (0.3)
SocAoG (our method)	**69.5 (0.8)**	**66.1 (0.7)**	**69.1 (0.5)**	**66.5 (0.8)**	60.1 (0.6)	**64.1 (0.8)**

Table 2: Performance comparison between BERT, BERT$_S$, GDPNet, SimpleRE, SocAoG$_{reduced}$, and SocAoG. We report 5-run average results and the standard deviation (σ).
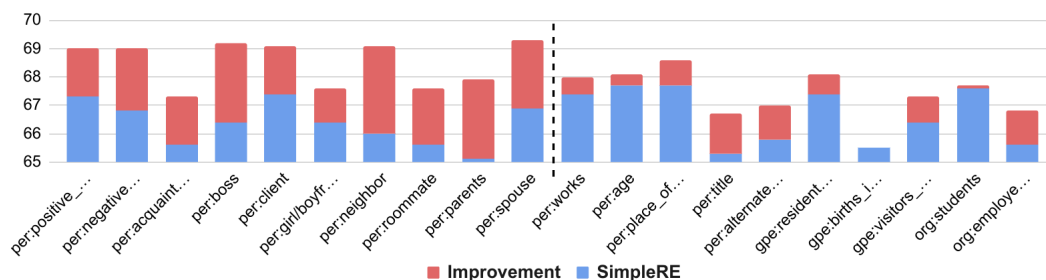

Figure 4: Performance boosts ($\mathbf{F}1$) of SocAoG compared to SimpleRE (Xue et al., 2020a) by relation type. The left bars to the dashed line are relations between humans, while the right ones are those between human and non-human entities.

64.1% *vs.* 63.2% for MovieGraph testing $\mathbf{F}1$. The results demonstrate our method can effectively take advantage of the attributes as cues for social relation predictions. We compare our SocAoG model with the existing model of highest accuracy (SimpleRE) by relation types, and see consistent improvements for all types. A part of the results are shown in Figure 4. We also observe that there are larger accuracy boosts for relations between human entities than non-human entities (*e.g.*, human-place), by an average of $+2.5\%$ *vs.* $+1.8\%$ in $\mathbf{F}1$, which is also reflected from Figure 4 (left 10 bars *vs.* right 10 bars). This can be explained as relation/attribute constraints are more meaningful for interpersonal relations, *e.g.*, there are more constraints for the relation between three humans than the relation between two humans and a place.

Table 2 also sees more accuracy improvement on MovieGraph dataset than DialogRE ($+3.2\%$ *vs.* $+6.0\%$ in test $\mathbf{F}1_c$ using SimpleRE as baseline). This is possibly because the dynamic inference nature of our method makes it effective for dealing with dialogues with more turns: while existing methods either truncate dialogues or use sliding windows, our method continuously updates the relation graph given an incoming turn. We case study the dynamic inference in the next subsection.

①	**S1, S2:**	Hi!
②	**S3:**	Hey!
③	**S4:**	So glad you came!
④	**S1:**	I can't believe Emma is already one!
⑤	**S2:**	I remember your first birthday! Ross was jealous of all the attention we were giving you. He pulled on his testicles so hard! We had to take him to the emergency room!
⑥	**S3:**	There's something you didn't know about your dad!
⑦	**S5:**	Hey Mr. and Mrs. Geller! Let me help you with that.
⑧	**S1:**	Thank you!
⑨	**S5:**	Oh man, this is great, uh? The three of us together again! You know what would be fun? If we gave this present to Emma from all of us!

Table 3: Dialogue example from the testing set of DialogRE (Yu et al., 2020).

5.5 Case Study on Dynamic Inference

Our method incrementally updates the relation and attribute information for a group of entities upon per utterance input with the proposed α–β–γ strategy. Such dynamic inference can potentially help reflect the evolving relations, unveil the reasoning process, and deal with long dialogues. Figure 5 shows the parse graph sequence by SocAoG inferring from a DialogRE testing dialogue as shown in Table 3. We can see that the method continuously refines the relation/attributes from an initial guess with incoming contexts, *e.g.* S2-S3: friends→parents in turn 5. Besides, the case also shows that attributes can aid relation predictions,
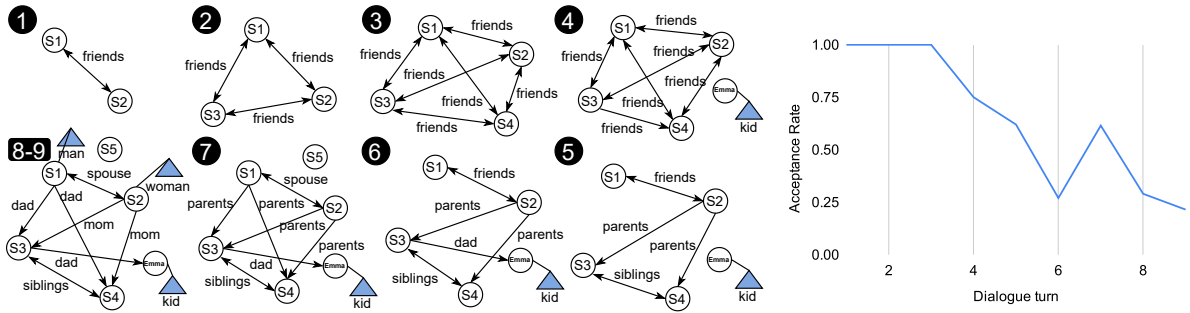
Figure 5: Left: inferred parse graph sequence from SocAoG based on the test dialogue in Table 3. Note that dad/mom are not distinguished in DialogRE. Right: model convergence measured by acceptance rate at each dialogue turn.

e.g., the inferred age of Emma clarifies her relation with S3. Moreover, since our method models the relation consistency among a group, it can predict the relation between two humans that do not talk directly. For example, S1 and S2 are inferred to be a couple by their dialogues with S5 in turn 7.

Figure 5 also plots the average MCMC acceptance rate for the case, as defined in Formula 10, indicting the convergence of the inference. We see that the algorithm only needs to update the current graph belief slightly with a new perceived utterance. A peak in the curve can indicate that a key piece of information is detected that contradicts the existing belief: *e.g.,* there is a peak of convergence curve in turn 7, which corresponds to *"S5: Hey Mr. and Mrs. Geller!"*, indicating that S1 and S2 are a couple rather than friends. As such, we can see the algorithm get several relations updated accordingly. We also show the convergence plots for 50 random testing cases from DialogRE in Figure 6, and the mean/standard deviation convergence rate as the black line/blue shade. We prove that our updating algorithm is robust for the converged results.
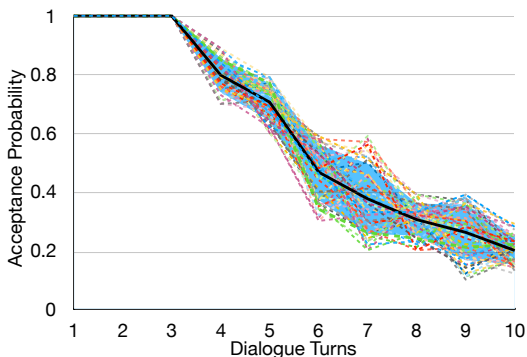

Figure 6: MCMC acceptance rate of the incremental parsing process. Dotted lines, black line, and blue shade are for samples, mean, and standard deviation, respectively.

Processes			$\mathbf{F}1(\sigma)$	$\mathbf{F}1_c(\sigma)$
α	β	γ		
✓			67.1 (0.5)	64.2 (1.1)
✓	✓		68.4 (0.8)	65.3 (0.6)
✓		✓	68.3 (0.4)	65.2 (0.7)
✓	✓	✓	**69.1 (0.5)**	**66.5 (0.8)**

Table 4: An ablation study on our parsing algorithm.

5.6 Ablation Study on α–β–γ

The α–β–γ strategy is designed to update relations and attributes jointly, having the input information flowing through the parse graph for the consistency of predictions. To validate the design, we ablate the processes on DialogRE to evaluate their impact on performance. Table 4 shows that α process, which is the discriminative model, makes the fundamental contribution, whereas β and γ processes alone cannot recognize social relations since they cannot perceive information from dialogues. Significantly, removing either one of the two processes will decrease the overall performance since the inference efficiency is reduced.

6 Conclusion

The paper proposes a SocAoG model with α–β–γ processes for the consistent inference of social relations in dialogues. The model can also leverage attribute information to assist the inference. MCMC is proposed to parse the relation graph incrementally, enabling the dynamic inference upon any incoming utterance. Experiments show that our model outperforms state-of-the-art methods; case studies and ablation studies are provided for analysis. In the future, we will further explore how different initialization of the parse graph could help warm start the inference under various situations and how multi-modal cues could be leveraged.

665

Acknowledgments

Ethical Considerations

Endowing AI to understand social relations is an essential step towards building emotionally intelligent agents. By jointly inferring individual attributes and social relations, our incremental parsing algorithm enables consistent and dynamic relational inference in dialogue systems, which can be remarkably useful for a wide range of applications such as a chatbot that constantly perceives new information and conducts social relation inference.

However, we never forget the other side of the coin. We emphasize that an ethical design principle must be in place throughout all stages of the development and evaluation. First, as discussed in Larson (2017), we model the attributes as a social construct from a performative view. For example, "gender performativity is not merely performance, but rather performances that correspond to, or are constrained by, norms or conventions and simultaneously reinforce them. Second, our model relies upon the attribute-category ascription provided by MovieGraph (Vicol et al., 2018) and Friends Central in Fandom. However, we acknowledge that the annotation could be prone to a partial understanding of human relationships, and the real situation could be more complicated. Lastly, self-identification should be the gold standard for ascribing attribute categories. Practitioners are suggested to prompt users to provide self-identification and respect the difficulties of respondents when asking. Our model helps increase the interpretability of the relational inference process by tracking the attributes and updating the relational belief. We expect that the biases from relation recognition can be easier to measure, and our α–β–γ processes may provide a multidimensional way for correcting them.

References

Christoph Alt, Marc Hübner, and Leonhard Hennig. 2019. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1388–1398, Florence, Italy. Association for Computational Linguistics.

Nguyen Bach and Sameer Badaskar. 2007. A review of relation extraction. *Literature review for Language and Statistics II*, 2:1–15.

Punam Bedi and Chhavi Sharma. 2016. Community detection in social networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(3):115–135.

David Belanger and Andrew McCallum. 2016. Structured prediction energy networks. In *International Conference on Machine Learning*, pages 983–992. PMLR.

Snigdha Chaturvedi, Mohit Iyyer, and Hal Daume III. 2017. Unsupervised learning of evolving relationships between literary characters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Hui Chen, Pengfei Hong, Wei Han, Navonil Majumder, and Soujanya Poria. 2020. Dialogue relation extraction with document-level heterogeneous graph attention networks. *arXiv preprint arXiv:2009.05092*.

Siddhartha Chib and Edward Greenberg. 1995. Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Patrick Doreian and Frans N Stokman. 1997. The dynamics and evolution of social networks. *Evolution of social networks*, 1(1).

Nan Du, Bin Wu, Xin Pei, Bai Wang, and Liutong Xu. 2007. Community detection in large-scale social networks. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 16–25.

Matthew R. Gormley, Mo Yu, and Mark Dredze. 2015. Improved relation extraction with feature-rich compositional embedding models. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1774–1784, Lisbon, Portugal. Association for Computational Linguistics.

Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former Friends: Unsupervised

learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, California. Association for Computational Linguistics.

Zhijing Jin, Yongyi Yang, Xipeng Qiu, and Zheng Zhang. 2020. Relation of the relations: A new paradigm of the relation extraction problem. *arXiv preprint arXiv:2006.03719*.

Nanda Kambhatla. 2004. Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 178–181, Barcelona, Spain. Association for Computational Linguistics.

Evgeny Kim and Roman Klinger. 2019. Frowning Frodo, wincing Leia, and a seriously great friendship: Learning to classify emotional relationships of fictional characters. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 647–653, Minneapolis, Minnesota. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Anna Kukleva, Makarand Tapaswi, and Ivan Laptev. 2020. Learning interactions and relationships between movie characters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9849–9858.

Shantanu Kumar. 2017. A survey of deep learning methods for relation extraction. *arXiv preprint arXiv:1705.03645*.

Brian Larson. 2017. Gender as a variable in natural-language processing: Ethical considerations. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 1–11, Valencia, Spain. Association for Computational Linguistics.

Zuchao Li, Hai Zhao, Rui Wang, and Kevin Parnow. 2020. High-order semantic role labeling. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1134–1151, Online. Association for Computational Linguistics.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. 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*, pages 1003–1011, Suntec, Singapore. Association for Computational Linguistics.

Makoto Miwa and Yutaka Sasaki. 2014. Modeling joint entity and relation extraction with table representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1858–1869, Doha, Qatar. Association for Computational Linguistics.

Raymond J Mooney and Razvan C Bunescu. 2006. Subsequence kernels for relation extraction. In *Advances in neural information processing systems*, pages 171–178.

Yannis Papanikolaou, Ian Roberts, and Andrea Pierleoni. 2019. Deep bidirectional transformers for relation extraction without supervision. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*, pages 67–75, Hong Kong, China. Association for Computational Linguistics.

Siyuan Qi, Yixin Zhu, Siyuan Huang, Chenfanfu Jiang, and Song-Chun Zhu. 2018. Human-centric indoor scene synthesis using stochastic grammar. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5899–5908.

Liang Qiu, Yizhou Zhao, Weiyan Shi, Yuan Liang, Feng Shi, Tao Yuan, Zhou Yu, and Song-Chun Zhu. 2020. Structured attention for unsupervised dialogue structure induction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1889–1899, Online. Association for Computational Linguistics.

Meng Qu, Tianyu Gao, Louis-Pascal Xhonneux, and Jian Tang. 2020. Few-shot relation extraction via bayesian meta-learning on relation graphs. In *International Conference on Machine Learning*, pages 7867–7876. PMLR.

Tianmin Shu, Michael S Ryoo, and Song-Chun Zhu. 2016. Learning social affordance for human-robot interaction. *arXiv preprint arXiv:1604.03692*.

Shashank Srivastava, Snigdha Chaturvedi, and Tom Mitchell. 2016. Inferring interpersonal relations in narrative summaries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Piotr Sztompka. 2002. Socjologia. *Analiza społeczeństwa, Znak, Kraków*, page 324.

Paul Vicol, Makarand Tapaswi, Lluis Castrejon, and Sanja Fidler. 2018. Moviegraphs: Towards understanding human-centric situations from videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8581–8590.

Tianfu Wu and Song-Chun Zhu. 2011. A numerical study of the bottom-up and top-down inference processes in and-or graphs. *International journal of computer vision*, 93(2):226–252.

Fuzhao Xue, Aixin Sun, Hao Zhang, and Eng Siong Chng. 2020a. An embarrassingly simple model for dialogue relation extraction. *arXiv preprint arXiv:2012.13873*.

Fuzhao Xue, Aixin Sun, Hao Zhang, and Eng Siong Chng. 2020b. Gdpnet: Refining latent multi-view graph for relation extraction. *arXiv preprint arXiv:2012.06780.*

Xue Yongqiang, Gao Baojiao, and Gao Jianfeng. 1997. The theory of thermodynamics for chemical reactions in dispersed heterogeneous systems. *Journal of colloid and interface science*, 191(1):81–85.

Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. 2020. Dialogue-based relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940, Online. Association for Computational Linguistics.

Godandapani Zayaraz et al. 2015. Concept relation extraction using naïve bayes classifier for ontology-based question answering systems. *Journal of King Saud University-Computer and Information Sciences*, 27(1):13–24.

Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella. 2003. Kernel methods for relation extraction. *Journal of machine learning research*, 3(Feb):1083–1106.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.

Shubin Zhao and Ralph Grishman. 2005. Extracting relations with integrated information using kernel methods. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 419–426, Ann Arbor, Michigan. Association for Computational Linguistics.

Yizhou Zhao, Liang Qiu, Wensi Ai, Feng Shi, and Song-Chun Zhu. 2020. Vertical-horizontal structured attention for generating music with chords. *arXiv preprint arXiv:2011.09078.*

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 207–212, Berlin, Germany. Association for Computational Linguistics.

Song-Chun Zhu and David Mumford. 2007. *A stochastic grammar of images*. Now Publishers Inc.

Song Chun Zhu, Yingnian Wu, and David Mumford. 1998. Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *International Journal of Computer Vision*, 27(2):107–126.

A   Appendices

ID	Subject	Relation Type	Object	Inverse Relation
1	PER	per:positive_impression	NAME	
2	PER	per:negative_impression	NAME	
3	PER	per:acquaintance	NAME	per:acquaintance
4	PER	per:alumni	NAME	per:alumni
5	PER	per:boss	NAME	per:subordinate
6	PER	per:subordinate	NAME	per:boss
7	PER	per:client	NAME	
8	PER	per:dates	NAME	per:dates
9	PER	per:friends	NAME	per:friends
10	PER	per:girl/boyfriend	NAME	per:girl/boyfriend
11	PER	per:neighbor	NAME	per:neighbor
12	PER	per:roommate	NAME	per:roommate
13	PER	per:children	NAME	per:parents
14	PER	per:other_family	NAME	per:other_family
15	PER	per:parents	NAME	per:children
16	PER	per:siblings	NAME	per:siblings
17	PER	per:spouse	NAME	per:spouse
18	PER	per:place_of_residence	NAME	gpe:residents_of_place
19	PER	per:place_of_birth	NAME	gpe:births_in_place
20	PER	per:visited_place	NAME	gpe:visitors_of_place
21	PER	per:origin	NAME	
22	PER	per:employee_or_member_of	NAME	org:employees_or_members
23	PER	per:schools_attended	NAME	org:students
24	PER	per:works	NAME	
25	PER	per:age	VALUE	
26	PER	per:date_of_birth	VALUE	
27	PER	per:major	STRING	
28	PER	per:place_of_work	STRING	
29	PER	per:title	STRING	
30	PER	per:alternate_names	NAME/STRING	
31	PER	per:pet	NAME/STRING	
32	GPE	gpe:residents_of_place	NAME	per:place_of_residence
33	GPE	gpe:births_in_place	NAME	per:place_of_birth
34	GPE	gpe:visitors_of_place	NAME	per:visited_place
35	ORG	org:employees_or_members	NAME	per:employee_or_member_of
36	ORG	org:students	NAME	per:schools_attended
37	NAME	unanswerable	NAME/STRING/VALUE	

Table 5: Relation types in DialogRE.

attributes	gender	male, female
	age	adult, kid, young adult, teenager, senior, baby
	ethnicity	caucasian, asian, arab, south-asian, hispanic, african, native american, other, aboriginal, african-american
	profession	photographer, cab driver, priest, writer, receptionist, delivery man, yoga instructor, chef, bartender, waitress, tailor, parking attendant, student, professional, lawyer, teacher, businessman, secretary, model, prince, banker, court reporter, intern, police officer, child psychologist, doctor, salesman/woman, hustler, bull rider, worker, doctors, businessman/woman, nurse, barman, janitor, policeman, inspector, FDA agent, counselor, waiter, judge, magician, prostitute, doorman, elevator operator, hotel manager, maid, bellhop, saleswoman, salesman, politician, driver, usher, actress, actor, florist, pilot, flight attendant, film/tv producer, building manager, paramedic, federal agent, postal worker, comic book artist, singer, executive, hockey player, referee, waiter/waitress, ex-soldier, receptionist, mafia boss, mafia member, musician, drug lord, fruit vendor, barber, masseuse, mental patient, mental patient, bus driver, night guard, housewife, editor, gardener, publisher, builder, elf, security guard, security chief, pedicurist, professor of defense against the dark arts, wandmaker, wizard, caretaker, ghost, villain, Philadelphia Eagles fan, cowboys America fan, bookmaker, unemployed, high school principal, jobless, racists, nuclear physicist, surgeon, soldier, colonel, professor, engineer, military officer, technician, game show host, police, robber, waiter/waitress, hitman, actor/actress, criminal, boxer, drug dealer, restaurant host, impersonator, military, trainer, manager, housekeeper, veterinarian, sportsperson, sports coach, sports agent, accountant, personal assistant, nanny, reporter, tv host, cameraman, tv presenter, cashier, artist, chauffeur, video artist, private investigator, administrator, tennis instructor, professional tennis player, detective, ticket collector, director, medical workers, hospital orderly, pharmacist, security officer, dental assistant, dentist, drug addict, registered sex offender, fetish worker, customer support, policemen, CEO, babysitter, assistant, principal, guidance counselor, farmer, entertaining, domestic worker, fisherman, author, psychologist, security person, tv personality, zeppelin crewman, king/queen, knight, journalist, assistant, weatherman, show host, make-up artist, seller, agent, tv show host, makeup artist, treasure hunter, naval officer, steward, ship captain, ship designer, sailor, designer, carpenter, valet, bail bondsman, court bailiff, court clerk, blackjack dealer, movie star, casino owner, casino manager, art director, executive recruiter, sports editor, cowboy, cowboy employer, hacker, investment counselor, hairdresser, sports commentator, chemist, government rep, vicar, robot, hotline agent, cook, surrogate date, philosopher, architect, record store owner, movie reviewer, call operator, bride, dog sitter, newspaper employer, vet, insurance broker, union leader, tv reporter, senator, rancher, locksmith, district attorney, store owner, smuggler, insurance agent, video editor, bouncer, trainee, real estate agent, prison guard, tour guide, mobster
relations		sibling, parent, cousin, customer, friend, stranger, spouse, colleague, boss, would like to know, lover, mentor, engaged, knows by reputation, acquaintance, roommate, best friend, antagonist, employed by, business partner, student, classmate, patient, teacher, child, heard about, enemy, employer of, psychiatrist, doctor, collaborator, ex-lover, landlord, superior, supervisor, grandchild, divorced, sponsor, ex-boyfriend, neighbor, fan, close friend, sister/brother-in-law, uncle, host, employer, step-mother, foster-son, family friend, godfather, godson, brother-in-law, nanny, grandparent, aunt, aide, students, family, customers, classmates, alleged lover, trainer, slave, hostage, robber, owner, instructor, competitor, fiancee, aunt/uncle, mother-in-law, girlfriend, killer, babysitter, one-night stand, boyfriend, tenant, distant cousin, father-in-law, mistress, agent, replacement, argue about the relationship, lawyer, ex-spouse, ex-girlfriend/ex-boyfriend, niece/nephew, parent-in-law, guardian, operative system, couple, goddaughter, customer, ex-neighbor, worker, vet, apprentice, public official, nurse, supporter, interviewee, interviewer, supporters, ex-fiance, fiance

Table 6: Attribute and relation types in MovieGraph.