Learning Semantic Correspondences from Noisy Data-text Pairs by Local-to-Global Alignments

Feng Nie[1,3*] **Jinpeng Wang**[2] **Chin-Yew Lin**[2]
[1]Interactive Entertainment Group, Tencent Inc, Shenzhen, China
[2]Microsoft Research Asia [3]Sun Yat-Sen University
{fengniesysu,wjppku}@gmail.com
cyl@microsoft.com

Abstract

Learning semantic correspondences between structured input data (e.g., slot-value pairs) and associated texts is a core problem for many downstream NLP applications, e.g., data-to-text generation. Large-scale datasets recently proposed for generation contain loosely corresponding data text pairs, where part of spans in text cannot be aligned to its incomplete paired input. To learn semantic correspondences from such datasets, we propose a two-stage local-to-global alignment (L2GA) framework. First, a local model based on multi-instance learning is applied to build alignments for texts spans that can be directly grounded to its paired structured input. Then, a novel global model built upon a memory-guided conditional random field (CRF) layer aims to infer missing alignments for text spans not supported by paired incomplete inputs, where the memory is designed to leverage alignment clues provided by the local model to strengthen the global model. In this way, the local model and global model can work jointly to learn semantic correspondences in the same framework. Experimental results show that our proposed method can be generalized to both restaurant and computer domains and improve the alignment accuracy.

1 Introduction

Discovering semantic correspondences between structured input (e.g., a set of slot-value pairs) and associated descriptions is a central step for many downstream NLP tasks. For example, in data-to-text generation, these correspondences indicate which subset of input data is verbalized in the description texts (what to say) and how data are described in natural language (how to say) (Angeli et al., 2010; Perez-Beltrachini and Lapata, 2018).

Recent neural methods for data-to-text generation require using large-scale training corpus. These datasets are often automatically constructed from parallel structured data and descriptions in websites. Some of the collected instances inevitably contain noise where input structured data are semantically inequivalent with associated texts (Perez-Beltrachini and Gardent, 2017; Nie et al., 2019). Fig. 1 depicts an example. The slot-value pair `Rating:low` in the input meaning representation (MR) is contradicted with the text span *highly recommended* in descriptions, while the text span *restaurant* in descriptions should refer to a slot-value pair `EatType:restaurant`, which is not in the paired MR. Previous work (Barzilay and Lapata, 2005; Liang et al., 2009; Perez-Beltrachini and Lapata, 2018) aligns text spans in descriptions by only focusing on building semantic correspondences with the paired MR. However, in the scenario of semantically inequivalent MR-text pairs, due to the noise of input, there is not sufficient information to induce alignments for text spans such as *restaurant* by merely looking at its current input.

To tackle the challenge of inducing correspondences (i.e., alignments) from semantically inequivalent data text pairs, we propose a local-to-global Alignment (L2GA) framework. This framework is composed of a local and a global model. The *local* model grounds text spans in description texts with corresponding slot-value pairs in its paired MR (e.g., *low price range* is aligned with slot `Price` in Fig. 1). The *global* model then exploits dependencies among correspondences in the entire corpus, so that alignments for text

*Contribution during internship at Microsoft.

Proceedings of the 28th International Conference on Computational Linguistics, pages 3050–3059
Barcelona, Spain (Online), December 8-13, 2020

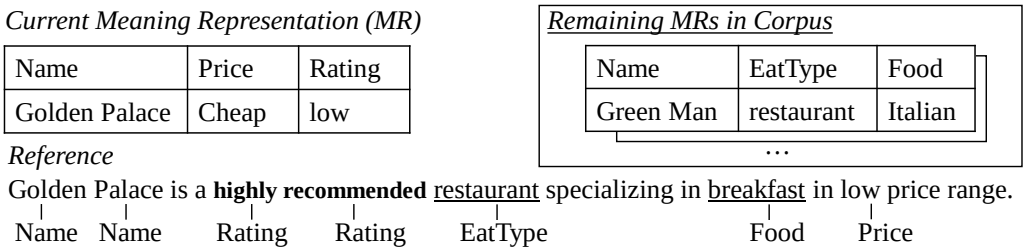

Figure 1: An example of semantically inequivalent MR-text pair. Underlined words refer to facts that are not supported by its paired incomplete MR but has close semantics with respect to other MRs. Text spans in **bold** refer to facts that are contradictory to its paired MR.

spans that are semantically relevant to some missing slot value pairs can be induced (e.g., *restaurant* is aligned with the slot `EatType` in Fig. 1).

Specially, our proposed L2GA breaks learning alignments into two steps. First, the *local* alignment model is a neural method optimized via a multi-instance learning paradigm (Perez-Beltrachini and Lapata, 2018) which automatically aligns each text span to a corresponding slot-value pair in its paired MR with the maximum semantic similarity.

The *global* alignment model is based on a memory-guided conditional random field (CRF) module, where *local* alignment model acts as an additional memory to guide CRF layer to jointly produce alignments for text spans in descriptions by leveraging information in the paired structured inputs as well as complementary information in the entire corpus simultaneously. In this way, the *global* alignment model is not only capable to induce alignments for text spans that are directly grounded to its paired MRs but also those text spans which cannot be aligned to their incomplete MRs with the help of global information. Note that alignments for training the CRF cannot be obtained directly, so we turn to leverage limited supervision provided in a data-text pair. Pseudo alignment labels are constructed using string matching heuristic between words and slots (e.g., *Golden Palace* is aligned with slot `Name` in Fig. 1), which left large portion of unmatched text spans (e.g., *low price* and *restaurant* cannot be directly matched in Fig. 1). In order to infer the alignments of unmatched words, we modify the calculation of sequence probability in the memory-guided CRF module by accumulating the probabilities over all possible sequential labels.

We conduct experiments on datasets in restaurant (Novikova et al., 2017a) and computer (Wang et al., 2017) domains to evaluate our proposed method. Experimental results show that our proposed method can improve the alignment accuracy and the effectiveness of introducing global alignment model for detecting noise presented in the original training corpus. In summary, our contributions are: 1) we propose to learn semantic correspondences for loosely corresponded data text pairs with both *local* and *global* supervisions; 2) we propose a local-to-global framework which not only induces semantic correspondences for words that are related to its paired input but also infers potential labels for text spans that are not supported by its incomplete input; 3) experimental results show that our proposed method can be generalized to both restaurant and computer domains and improve the alignment accuracy.

2 Task Formulation

Here, we provide a brief description of learning alignments in semantically inequivalent data-text pairs. Given a corpus with paired meaning representations (MR) and text descriptions $\{(R, X)\}_{i=1}^{N}$. The input MR $R = (r_1, \ldots, r_M)$ is a set of slot-value pairs $r_j = (s_j, v_j)$, where each r_j contains a slot s_j (e.g., `Price`) and a value v_j (e.g., `Cheap`). The corresponding description $X = (x_1, \ldots, x_T)$ is a sequence of words describing the MR. The task is to match every word x_i in text X with a possible slot, where the slot can be one of M slots presented its paired MR or one of K slots in the entire corpus, where $K \geq M$. Note that words such as stopwords in the description texts are irrelevant to any slot, we add a special label `NULL` indicating words without corresponding alignments. An example of alignments in a data-text pair is shown in Fig. 1. Note that the example depicts alignments for text spans that cannot be grounded by its paired MR in two scenarios: (1) input MR contains errors where slot-value pairs are contradictory to texts (e.g., slot-value pair `Rating:low` is contradictory to text span *highly recommended*), (2) extra slots in

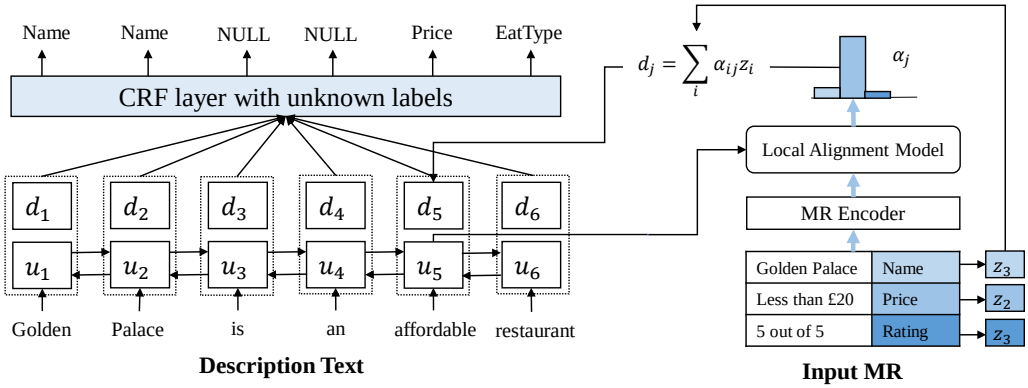

Figure 2: An illustration of Local-to-Global Alignment (L2GA). It consists of a *local* alignment model and a *global* alignment model built on a CRF layer.

descriptions (e.g., `EatType : restaurant`). In the next section, we present our approach to address this task.

3 Approach

Our proposed method is a local-to-global alignment (L2GA) model, as shown in Fig. 2. It is a two-stage alignment framework. The *local* model first encodes both description X and its paired MR R using contextualized encoders, then aligns each word in description X with a most relevant slot-value pair in MR R by computing semantic similarity based on the contextualized encodings. For semantically inequivalent data-text pairs, input MRs are not guaranteed to cover all information of corresponding descriptions. To tackle the challenge, a *global* model with a memory guided CRF layer is proposed to exploit dependencies among alignments over the entire corpus and therefore produces missing alignments for text spans not existed in the paired MR. Note that the memory is designed to integrate the alignments captured by the *local* model to improve the *global* model.

3.1 Local Alignment Model

The local model tries to induce alignments for words in description texts with respect to its paired input MR. Given a data-text pair (R, X), we assume that words in the description X are positively related for some slot-value pairs in R but we do not know which ones. Following (Perez-Beltrachini and Lapata, 2018), we formulate this task into a multi-instance learning problem (Keeler and Rumelhart, 1992), where fine-grained annotations (i.e., alignments for text spans) are discovered from the coarse level supervisions (i.e., the similarities of MR-text pairs). We first introduce the encoders for input MR R and description X, then the alignment objectives to acquire the alignments for words in text X.

MR Encoder: A slot-value pair r in MR can be treated as a short sequence $w_1, \ldots, w_n$ by concatenating words in its slot and value. The word sequence is first represented as a sequence of word embedding vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ using a pre-trained word embedding matrix $\mathrm{E_w}$, and then passed through a bidirectional LSTM layer to yield the contextualized representations $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_n)$. To produce a summary context vector over variable-length sequences, we adopt the same self-attention structure in (Zhong et al., 2018) to obtain the vector of slot-value pair $\mathbf{c}$.

$$\mathbf{c} = \sum_i \beta_i \mathbf{h}_i; \quad \beta = \mathrm{softmax}(\mathrm{W_s}\mathbf{H}) \tag{1}$$

where W_s is a trainable parameter $\mathrm{W_s}$ and β is the learned importance. We also embed each slot s_i into a slot vector as

$$\mathbf{z}_i = \mathrm{E_z}(\mathrm{s_i}) \tag{2}$$

where E_z is a trainable slot embedding matrix.

Sentence Encoder: For description $X = (x_1, \ldots, x_T)$, each word x_t is first embeded into vector $\mathbf{e}_t$ by concatenating the word embedding and character-level representation generated with a group of convolutional neural network (CNNs). Then we feed the word vectors $\mathbf{e}_1, ..., \mathbf{e}_T$ to a bidirectional LSTM to obtain contextualized vectors $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_T)$.

$$\mathbf{u}_t = \mathrm{biLSTM}_{\mathrm{sent}}(\mathbf{e}_{\mathrm{t}}, \mathbf{u}_{\mathrm{t}-1}) \tag{3}$$

Alignment Objective: Our goal is to maximize the similarity score between MR R and description X. The similarity R and X is in turn defined on the top of the similarity scores among context vector $\mathbf{c}_1, ..., \mathbf{c}_M$ of slot-value pairs in MR and sentence vector $\mathbf{u}_1, \ldots, \mathbf{u}_T$ of description produced by MR encoder and sentence encoder respectively.

$$S_{(R,X)} = \frac{1}{T} \sum_{t=1}^{T} \max_{i \in \{1,\ldots,M\}} \mathbf{c}_i \cdot \mathbf{u}_t \tag{4}$$

where $\cdot$ refers inner product of two vectors. The function in Eq. 4 aims to align each word with the best scoring slot-value pair. Note that each word x_t is aligned with a slot-value pair r_i if the similarity (i.e., inner product between two vectors) is larger than a threshold. To train the local alignment model, the following loss function encourages related MR R and description X to achieve higher similarity than other MR $R^{'} \neq R$ and texts $X^{'} \neq X$ by a margin γ:

$$\begin{aligned} L_{co} = & \max(0, \gamma - (S_{(R,X)} - S_{(R,X')})) \\ & + \max(0, \gamma - (S_{(R,X)} - S_{(R',X)})) \end{aligned} \tag{5}$$

3.2 Global Alignment Model

Since the data-text pairs are semantically inequivalent, part of text spans are not supported by the noisy paired input. To induce alignments for those text spans, our proposed *global* alignment model is based on a CRF module which is capable of leveraging dependencies among alignments. Compared with the conventional sequence labeling, our scenario differs in two aspects. First, we employ a weak supervision method to obtain the training labels for sequence labeling. Second, the alignment supervision provided by the *local* model is leveraged for sequence labeling.

To provide labels for training a CRF module, we first generate pseudo labels for words in texts by exact string matching, where conflicted matches are resolved by maximizing the total number of matched tokens (Shang et al., 2018). Based on the result of dictionary matching, each word falls into one of three categories: 1) it belongs to an entity mention with one slot in its paired MR; 2) it belongs to an (unknown) entity where its slot is either not directly labeled using string matching or not represented in its paired MR; 3) it is marked as a non-entity[1]. Therefore, we change the sequence paths in CRF layer to allow inducing alignments for words with unknown types. Moreover, *local* model provides alignment clues for text spans. Some of them are ignored by string heuristics but are semantically relevant to one slot in paired MRs (e.g., *affordable* is relevant to `Price` in Fig. 2). These alignments are treated as a soft memory to guide the CRF layer.

Modified LSTM-CRF: In conventional LSTM-CRF based sequence labeling model (Lample et al., 2016), given the text description $X = \{x_t\}_{t=1}^{T}$ and the pseudo labels $Y = \{y_t\}_{t=1}^{T}$. We first obtain contextual representations $\mathbf{U}$ for words in description X using the Eq. 3, and context vector $\mathbf{u}_t$ for word x_t is decoded by a linear layer $\mathrm{W_c}$ into the label space to compute the score P_{t,y_t} for label y_t.

$$P_t = \mathrm{W_c}\mathbf{u}_{\mathrm{t}} \tag{6}$$

On top of the model, a CRF (Lafferty et al., 2001) layer is applied to capture the dependencies among predicted labels. We define the score of the predicted sequence $(y_1, ..., y_T)$ as:

$$s(X, Y) = \sum_{t=0}^{T} \Phi_{y_t, y_t+1} + \sum_{t=1}^{T} P_{t,y_t} \tag{7}$$

[1]Note that all stopwords without exact string matching are treated as a non-entity, where we assign a special label `NULL`.

where Φ_{y_t,y_t+1} is the transition probability from a label y_t to its next label y_{t+1}. Φ is a $(K+2)\times(K+2)$ matrix, where K is the number of distinct labels (i.e., unique slots in the entire corpus). Two additional labels `start` and `end` are used (only used in the CRF layer) to represent the beginning and end of a sequence, respectively.

The conventional CRF layer maximizes the probability of sequences using the provided labels. However, there are unmatched text spans with a possible label in our scenario (e.g., text spans *restaurant* and *affordable* in Fig. 2). We instead compute probability for sequence Y by enumerating all possible tags for unmatched text spans. The optimization goal is defined as:

$$p(Y|X) = \frac{\sum_{\hat{Y}\in Y_{possible}} e^{s(X,\hat{Y})}}{\sum_{\hat{y}\in Y_X} e^{s(X,\hat{Y})}} \tag{8}$$

where Y_X refers to all the possible sequential labels for X, and $Y_{possible}$ refer to sequential labels obtained by enumerating all possible alignments for unmatched text spans. Note that, if there are no unmatched text spans in description text X, it is equivalent to the conventional CRF.

Integrate Local Alignment Clues: The *local* alignment model can provide alignment supervisions for text spans that are lexically different but semantically relevant to slot-value pairs in paired MRs. To incorporate the induced semantic correspondences provided by *local* alignment model, we design a specific memory to guide the CRF module. Specially, for each word x_t in description X, we select a most probable slot s_i using the similarity score calculated in Eq. 4, and compute the slot representation $\mathbf{d}_t$ as follows

$$\mathbf{d}_t = \sum_{i=1}^{n} \alpha_{t,i}\mathbf{z}_i; \quad \alpha_{t,i} = \text{softmax}(\mathbf{c}_i \cdot \mathbf{u}_t) \tag{9}$$

where $\alpha_{t,i}$ refers to the probability that word x_t is related to slot s_i in MR and $\mathbf{z}_i$ is the slot embedding for slot s_i defined in Eq. 2. We then utilize the alignment information $\mathbf{d}_t$ to help the calculation of the prediction score P_t in Eq. 6. Concretely, we modified the Eq. 6 as following:

$$P_t = \mathrm{W_c}[\mathbf{u_t}, \mathbf{d_t}] \tag{10}$$

where $[,]$ refers to concatenation of two vectors. In this way, alignments produced by the local alignment model can act as a guidance to help inducing labels of entities in texts.

3.3 Model Training and Inference

During training, we optimize the global model by minimizing negative log-likelihood $p(Y|X)$ of the score defined in Eq. 8. We optimize the *local* and *global* model jointly using the following training loss:

$$L = -\log p(Y|X) + L_{co} \tag{11}$$

where L_{co} is the alignment objective of *local* alignment model defined in Eq. 5. For inference, we apply Viterbi decoding to obtain the alignments for description texts by maximizing the score defined in Eq. 7.

4 Experiments

4.1 Experimental Setup

Data: we conduct experiments on two data-to-text datasets. (1) **E2E** (Novikova et al., 2017b): E2E challenge is a dataset in restaurant domain. It has 42,061, 4,672 and 4,693 MR-text pairs for training, validation and testing, respectively. Note that every input MR in this dataset has 8.65 different references on average. The dataset contains 8 unique slots in the entire dataset (i.e., Name, Near, EatType, Rating, Food, Price, Area, FamilyFriendly). This dataset contains roughly 18% loosely aligned data-text pairs reported in (Novikova et al., 2017b). (2) **Computer** (Wang et al., 2017): It is a dataset collected from 'Computers & Tablets' in Amazon.com. The original dataset contains incoherent descriptions which

	E2E						Computer					
	Full			Noisy Data-text Pairs			Full			Noisy Data-text Pairs		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
MIL	83.21	83.48	83.34	68.31	71.93	70.07	62.82	35.76	45.58	62.35	34.44	44.37
Distant LSTM-CRF	72.44	74.03	73.23	59.77	66.75	63.07	**89.32**	40.81	56.03	**89.19**	38.01	53.31
Modified LSTM-CRF	89.78	94.84	92.24	74.13	90.29	81.41	55.54	71.55	62.54	54.42	70.25	61.33
L2GA	**92.93**	**96.69**	**94.77**	**76.45**	**91.98**	**83.49**	58.85	**72.86**	**65.11**	57.19	**71.33**	**63.48**

Table 1: Alignment results of different methods on two datasets.

merely list attributes in the input. We filter out these data-text pairs, and the remaining 17,404 data-text pairs have been divided into three parts to provide training (70%), validation (10%) and test sets (20%). The filtered dataset contains 43 unique slots, and has 133.4 words in textual description on average[2]. This dataset contains roughly 70% loosely aligned data-text pairs[3].

Evaluation: It is difficult to evaluate the accuracy of alignments for the entire corpus, since the alignments are not provided in the original data. For each dataset, we recruited three annotators who are familiar with the corresponding domain to label the testsets. For **E2E** dataset, the testset contains 630 unique MR. We randomly sample a reference for each MR and the annotators are asked to label 630 data-text pairs. For **Computer** dataset, we randomly sample 100 data-text pairs from the testset, which includes 480 sentences. There exists a fraction of 7.3% and 25.7% where agreements not made in **E2E** and **Computer** dataset respectively, and the fourth annotator is invited to make the final decisions for these conflicts. We report the precision (P), recall (R) and F1 accuracy on the annotated testsets for evaluation.

Training Configuration: For all models, we use pre-trained GloVe vectors (Pennington et al., 2014) and character embeddings (Hashimoto et al., 2017) for initialization of word embeddings. Model parameters were tuned on the development set. We use the stochastic gradient descent with an initial learning rate 0.015, and decay by 0.9 for every epoch after 5 epochs. The dimensions of trainable hidden units in LSTMs are 400 and 600 for **E2E** and **Computer** datasets respectively. The gradient is truncated by 5. The hyper-parameter γ in Eq. 5 is set to 0.1. For model L2GA, the *local* alignment model is trained by 5 epochs in advance.

Baselines: We compare our proposed alignment model with the following neural baselines: i) MIL (Perez-Beltrachini and Lapata, 2018), which refers to the *local* model. Note that each word is assigned to a slot if the semantic similarity defined in Eq. 4 is larger than 0.1; ii) Distant LSTM-CRF (Giannakopoulos et al., 2017), which is a dictionary based sequence labeling model for distant supervised name entity recognition (NER). We make adaptation by treating the paired MR as the dictionary to create initial training labels described in Section 3.2 and train a LSTM-CRF model based on the pseudo training data; iii) Modified LSTM-CRF, which is our proposed *global* model without leveraging *local* alignment information as described in Section 3.2.

4.2 Results

Table 1 presents the results of our proposed method (L2GA) with other baselines. L2GA achieves best F1 scores in two datasets compared with both *local* and variations of *global* models.

MIL is the *local* alignment model, which induces alignments for each text span by selecting the most relevant slot from the paired MR. It is incapable of producing potential alignments for text spans that are not supported by its paired MR. As a result, MIL performs worst in **Computer** dataset, where large portion of data-text pairs are noisy. L2GA leverages dependencies of alignments globally using a specified CRF module and achieves 11.43% and 19.53% F1 improvements over `MIL` on **E2E** and **Computer** datasets respectively. The results demonstrate the effectiveness of introducing *global* alignment model for semantically inequivalent data-text pairs.

[2]Note that the original dataset contains ambiguous and irrelevant slots. We filter out this irrelevant slots in advance and we will release this filtered dataset on the acceptance of this paper.

[3]We manually annotate 100 data-text pairs and 71 of description texts contain missing slot information with respect to its paired MR.

	E2E			Computer		
	P (%)	R (%)	F1 (%)	P (%)	R (%)	F1 (%)
L2GA	92.93	96.69	94.77	58.85	72.86	65.11
Local CRF	85.58	88.86	87.24	75.39	46.33	57.39

Table 2: Different combinations of local models with sequence labeling framework

	BLEU (%)	Err. (%)
S2S	66.15	59.37 (374/630)
S2S+L2GA	65.21	13.33 (84/630)
(Dusek et al., 2019)	65.87	9.70 (62/630)

Table 3: Generation results trained on different training corpus in **E2E** dataset.

The other two methods are distant supervised sequence labeling approaches, which can be treated as two variations of our proposed *global* alignment model. The Distant LSTM-CRF applies a LSTM-CRF framework directly on the training labels generated by string matching. It ignores the alignments for unmatched text spans, and yields really low recalls in two datasets. While Modified LSTM-CRF is able to infer alignments for unmatched text spans by changing the sequence probabilities and therefore, achieves higher F1 scores compared with Distant LSTM-CRF method. The result illustrates the necessity of modeling sequence probabilities for unmatched text spans. However, above two methods only leverages the alignment information in data-text pairs during the construction of pseudo labels. Part of text spans that are semantically equivalent but lexically different to slot-value pairs in paired MRs can be ignored (e.g., *affordable* is closely related to the slot-value pair `Price:Cheap` in Fig. 2). Our proposed method L2GA leverages such alignment information by dynamically integrating the alignments learned in *local* model into CRF module and therefore achieves substantial improvements.

As our proposed method targets on learning the alignments for semantically inequivalent data-text pairs, we pick the data-text pairs from the testsets where the description texts contains additional or contradicted alignments compared to the original MR, and report the performance of each method on the *Noisy* data-text pairs[4]. Results in Table 1 show that the *local* alignment model MIL decreases dramatically in term of recall in **E2E** dataset, while methods with *global* alignments such as Modified LSTM-CRF and L2GA are less sensitive. As the **Computer** dataset with larger portion of semantically inequivalent data-text pairs, the *local* model MIL failed to learn alignments for most of text spans and produces the lowest recall compared to methods leveraging global information. The results further confirms that global alignment information is essential for learning alignments in semantically inequivalent data-text pairs. Moreover, L2GA achieves higher precision than Modified LSTM-CRF baseline in both settings. The improvements illustrates that integrating *local* alignment model can bring complementary alignment guidance for *global* methods.

4.3 Analysis

4.3.1 Effect of Dynamic Combination

We also investigate different ways of incorporating *local* model with the sequence labeling framework. A straight forward way is to train a LSTM-CRF model with the alignments produced by the *local* models, which referred as Local CRF in Table 2. L2GA outperforms Local CRF by dynamically integrating the alignment results provided by the *local* model and avoiding label noise for training, therefore achieves better result.

4.3.2 How can Alignment Help Generation?

In this section, we provide an extrinsic evaluation by testing whether alignments can be beneficial to neural generation. Inspired by recent work (Dusek et al., 2019) on removing semantically inequivalent

[4] 127 out of 630 in **E2E** and 71 out of 100 in **Computer** testsets are recognized as semantically inequivalent data-text pairs.

Meaning Representation (MR)

Name	Rating	EatType	Price	Near	Food	Area	FamilyFriendly
The Cricketers	high	restaurant	£20-25	All Bar One	English	riverside	no

MIL:	**Name** **Food** **Near** The Cricketers is a kid friendly restaurant that serves English food near All Bar One in the **Area** **Price** **Rating** riverside area . It has a price range of 20-25 pounds and is a highly rated restaurant .
Modified LSTM-CRF:	**Name** **FamilyFriendly** **EatType** **Food** **Near** The Cricketers is a kid friendly restaurant that serves English food near All Bar One in the **Area** **Price** **Rating** **EatType** riverside area . It has a price range of 20-25 pounds and is a highly rated restaurant .
L2GA:	**Name** **FamilyFriendly** **EatType** **Food** **Near** The Cricketers is a kid friendly restaurant that serves English food near All Bar One in the **Area** **Price** **Rating** **EatType** riverside area . It has a price range of 20-25 pounds and is a highly rated restaurant .

Figure 3: An example of alignments produced by different models. Text spans with color indicate recognized entities, and texts above the highlighted spans refer to the corresponding labels.

	Total	Name	Near	EatType	Rating	Food	Price	Area	FamilyFriendly
L2GA	**94.77**	99.84	**99.67**	**96.97**	**86.76**	**92.87**	89.65	92.82	93.07
MIL	83.34	96.33	98.53	51.45	74.05	91.66	**89.77**	69.13	**94.24**
Modified LSTM-CRF	92.24	**100.00**	99.62	96.89	85.11	92.47	87.90	**92.89**	75.04

Table 4: Detailed alignment results of different models over each slot in **E2E** dataset.

noise in training corpus, it is viable to produce a refined MR with the learned alignments for description texts. Specifically, slot-value pairs (e.g., EatType:restaurant) are recovered by using the text spans (e.g., restaurant) with its corresponding alignments (e.g., *EatType*). For slots with string values (e.g., *Name*), the slot-value pairs (e.g., Name:Golden Palace) are recovered by recognizing text spans (e.g., Golden Palace) as slot values. For slots with categorical values (e.g., *Price*), we retrieve a most frequent slot-value pair (e.g., Price:high) in the training corpus with the same text span (e.g., expensive). In this way, a refined training corpus is produced. We use the new training corpus to train a sequence-to-sequence (S2S) generation model. To evaluate the correctness of generation, a well-crafted rule-based aligner built by (Juraska et al., 2018) is adopted to approximately reflect the semantic correctness. The error rate is calculated by matching the slot values in output texts containing missing or conflict slots in the realization given its input MR. The generation results are shown in Table 3. Vanilla S2S model trained on semantically inequivalent data-text pairs performs poorly in semantic correctness. After training on the corpus refined by our proposed L2GA method, S2S model can reduce the inconsistent errors in a large margin. Note that (Dusek et al., 2019) clean noise in training corpus by rich heuristic rules, L2GA refines the training corpus by learned alignments and performs competitively.

4.4 Qualitative Analysis

Fig. 3 gives the alignment results produced by different models. We can see that *local* model MIL cannot induce the alignment for the text span *kid friendly* as it is contradicted with the slot-value pair `FamilyFriendly:no`. While *global* models can induce the semantic correspondence for the text span *kid friendly* with the corresponding label `FamilyFriendly`. Moreover, the Modified LSTM-CRF has difficulty in labeling lexically different but semantically equivalent text span *highly*. While L2GA can dynamically integrate the alignment results provided by the *local* model, therefore produce the correct alignment `Rating` for the text span *highly rated* correctly.

5 Related Work

Previous work exploiting loosely aligned data and text corpora have mostly focused on discovering verbalisation spans for data units. These line of work usually follows a two stage paradigm, where data

units are first aligned with sentences from related corpora using heuristics and then subsequently extra content is discarded in order to retain only text spans verbalising the data. (Belz and Kow, 2010) use a measure of association between data units and words to obtain verbalisation spans. (Walter et al., 2013) extract patterns from paths in dependency trees. One exception is (Perez-Beltrachini and Lapata, 2018), the induced alignments are used to guide the generation. Our work takes a step further to also induce alignments for text spans not supported by the noisy paired input with possible semantics.

Our work is also related to previous work on extracting information from user queries with the backend data structure. Most of these approaches contain two steps. Initially, a separate model is applied to match the unstructured texts with relevant input records and then an extraction model is learned based on collected annotations. (Agichtein and Ganti, 2004) and (Canisius and Sporleder, 2007) train a language model on data records to identify related text spans in book description. Several approaches train a CRF based extractor to detect the related text spans (Michelson and Knoblock, 2008; Li et al., 2009). (Bellare and McCallum, 2009) apply a generalized expectation criteria to learn alignments between database and the texts, and train the information extractor to induce semantic annotations for text spans. Compared to these work, our approach is an unified neural based alignment model which avoids the error propagation of each step.

6 Conclusion

In this paper, we study the problem of learning alignments in semantically inequivalent data-text pairs. We propose a local-to-global framework which not only induces semantic correspondences for words that are related to its paired input but also infers potential labels for text spans that are not supported by its incomplete input. Our proposed method generalizes to both restaurant and computer domains and improves the alignment accuracy. In the future, we will explore pre-trained language models (Devlin et al., 2019) for contextualized encoders to improve the alignment accuracy.

7 Acknowledgement

We would like to thank Zhirui Zhang, Lijun Wu, Hui Liu, Jiacheng Gu and the anonymous reviewers for their constructive feedback and useful comments. This work was completed in partial fulfillment for the PhD degree of the first author.

References

Eugene Agichtein and Venkatesh Ganti. 2004. Mining reference tables for automatic text segmentation. In *KDD*.

Gabor Angeli, Percy Liang, and Dan Klein. 2010. A simple domain-independent probabilistic approach to generation. In *EMNLP*, October.

Regina Barzilay and Mirella Lapata. 2005. Modeling local coherence: An entity-based approach. In *ACL*.

Kedar Bellare and Andrew McCallum. 2009. Generalized expectation criteria for bootstrapping extractors using record-text alignment. In *EMNLP*.

Anja Belz and Eric Kow. 2010. Extracting parallel fragments from comparable corpora for data-to-text generation. In *INLG*.

Sander Canisius and Caroline Sporleder. 2007. Bootstrapping information extraction from field books. In *EMNLP-CoNLL*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL)*, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Ondrej Dusek, David M. Howcroft, and Verena Rieser. 2019. Semantic noise matters for neural natural language generation. *CoRR*, abs/1911.03905.

Athanasios Giannakopoulos, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl. 2017. Unsupervised aspect term extraction with b-LSTM & CRF using automatically labelled datasets. In *WASSA*, Copenhagen, Denmark, September. Association for Computational Linguistics.

Kazuma Hashimoto, Caiming Xiong, Yoshimasa Tsuruoka, and Richard Socher. 2017. A joint many-task model: Growing a neural network for multiple NLP tasks. In *EMNLP*, Copenhagen, Denmark, September. Association for Computational Linguistics.

Juraj Juraska, Panagiotis Karagiannis, Kevin Bowden, and Marilyn Walker. 2018. A deep ensemble model with slot alignment for sequence-to-sequence natural language generation. In *NAACL*, pages 152–162, New Orleans, Louisiana, June. Association for Computational Linguistics.

Jim Keeler and David E. Rumelhart. 1992. A self-organizing integrated segmentation and recognition neural net. In J. E. Moody, S. J. Hanson, and R. P. Lippmann, editors, *NIPs*.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *NAACL*.

Xiao Li, Ye-Yi Wang, and Alex Acero. 2009. Extracting structured information from user queries with semi-supervised conditional random fields. In *SIGIR*.

Percy Liang, Michael I. Jordan, and Dan Klein. 2009. Learning semantic correspondences with less supervision. In *ACL*.

Matthew Michelson and Craig A. Knoblock. 2008. Creating relational data from unstructured and ungrammatical data sources. *J. Artif. Intell. Res.*, 31.

Feng Nie, Jin-Ge Yao, Jinpeng Wang, Rong Pan, and Chin-Yew Lin. 2019. A simple recipe towards reducing hallucination in neural surface realisation. In *ACL*, Florence, Italy, July. Association for Computational Linguistics.

Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry, and Verena Rieser. 2017a. Why we need new evaluation metrics for NLG. In *EMNLP*, Copenhagen, Denmark, September. Association for Computational Linguistics.

Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. 2017b. The E2E dataset: New challenges for end-to-end generation. In *SIGDAL*, Saarbrücken, Germany, August. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*, Doha, Qatar, October. Association for Computational Linguistics.

Laura Perez-Beltrachini and Claire Gardent. 2017. Analysing data-to-text generation benchmarks. In *INLG*, Santiago de Compostela, Spain, September.

Laura Perez-Beltrachini and Mirella Lapata. 2018. Bootstrapping generators from noisy data. In *ACL*. Association for Computational Linguistics, June.

Jingbo Shang, Liyuan Liu, Xiaotao Gu, Xiang Ren, Teng Ren, and Jiawei Han. 2018. Learning named entity tagger using domain-specific dictionary. In *EMNLP*, Brussels, Belgium. Association for Computational Linguistics.

Sebastian Walter, Christina Unger, and Philipp Cimiano. 2013. A corpus-based approach for the induction of ontology lexica. In *INLG*.

Jinpeng Wang, Yutai Hou, Jing Liu, Yunbo Cao, and Chin-Yew Lin. 2017. A statistical framework for product description generation. In *IJCNLP*, Taipei, Taiwan, November.

Victor Zhong, Caiming Xiong, and Richard Socher. 2018. Global-locally self-attentive encoder for dialogue state tracking. In *ACL*, Melbourne, Australia, July. Association for Computational Linguistics.